PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

B6

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, C12Q 1/68, G01N 33/50, 33/53, 33/68, A61K 38/17

A1

(11) International Publication Number: WO 98/42738

(43) International Publication Date: 1 October 1998 (01.10.98)

(21) International Application Number: PCT/US98/05311

(22) International Filing Date: 19 March 1998 (19.03.98)

(30) Priority Data:

60/041,281	21 March 1997 (21.03.97)	US
60/041,276	21 March 1997 (21.03.97)	US
60/042,344	21 March 1997 (21.03.97)	US
60/041,277	21 March 1997 (21.03.97)	US
60/048,355	30 May 1997 (30.05.97)	US
60/048,096	30 May 1997 (30.05.97)	US
60/048,351	30 May 1997 (30.05.97)	US
60/048,154	30 May 1997 (30.05.97)	US
60/048,160	30 May 1997 (30.05.97)	US
60/048,069	30 May 1997 (30.05.97)	US
60/048,131	30 May 1997 (30.05.97)	US
60/048,186	30 May 1997 (30.05.97)	US
60/048,095	30 May 1997 (30.05.97)	US
60/048,187	30 May 1997 (30.05.97)	US
60/048,099	30 May 1997 (30.05.97)	US
60/050,937	30 May 1997 (30.05.97)	US
60/048,352	30 May 1997 (30.05.97)	US
60/048,135	30 May 1997 (30.05.97)	US
60/048,188	30 May 1997 (30.05.97)	US
60/048,094	30 May 1997 (30.05.97)	US
60/048,350	30 May 1997 (30.05.97)	US
60/054,804	5 August 1997 (05.08.97)	US
60/056,370	19 August 1997 (19.08.97)	US
60/060,862	2 October 1997 (02.10.97)	US

(71) Applicant (*for all designated States except US*): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Avenue, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown, MD 20874 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). DUAN, Roxanne [US/US]; 4541 Fairfield Drive, Bethesda, MD 20814 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 M. Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US).

(74) Agents: BROOKES, Anders, A. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

(54) Title: 87 HUMAN SECRETED PROTEINS

(57) Abstract

The present invention relates to 87 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

*(Referred to in PCT Gazette No. 46/1998, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

87 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoeitin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein , a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH_2PO_4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched , for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

The translation product of this gene shares sequence homology with nucleolin, which is thought to be important in macromolecule binding, as well as some membrane proteins. Preferred polypeptide fragments comprise the amino acid sequence: DPEAADSGEPQNKRTPDLPEEEYVKEEIQENEEAVKKMLVEATREFEEVVVDES (SEQ ID NO:239); QKLKRKAEEDPEAADSGEPQNKRTPDLPEEEYVKEEIQENEEAVKKMLVEATREFEEVVVDES (SEQ ID NO:240); KAMEKSSLTQHSWQSLKDRYLKHLRGQEHKYLLGDAPVSPSSQKLKRKAEEDPEAADSGEPQNKRTPDLPEEEYVKEEIQENEEAVKKMLVEATREFEEVVVDESPPDFEIHI (SEQ ID NO:241). Also preferred are the polynucleotide fragments encoding these polypeptide fragments.

This gene maps to chromosome 16, and therefore can be used as a marker in linkage analysis for chromosome 16.

This gene is expressed primarily in brain and kidney and to a lesser extent in wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cell-cell interaction or cell-matrix interaction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and kidney, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:125 as residues: Met-1 to Trp-10.

The tissue distribution and homology to nucleolin indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases involving cell-cell interaction or cell-extracellular matrix interaction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

The translation product of this gene shares sequence homology with a porcine zona pellucida protein ZPDS.1711. (See Accession No. R39356.) These two proteins have weak homology with *Drosophila* commissureless and metal homeostasis proteins which are thought to be important in controlling growth cone guidance across the CNS midline and protecting cells against reactive oxygen toxicity. thus, based on homology, it is likely that this gene also be involved in development. Preferred polypeptide fragments comprise the amino acid sequence: LPSYDEAERTKAEATIPLVPGRDEDF VGRDDFDDADQLRIGNDGIFMLTFFMAFLFNWIGFFLSFCLTTSAAGRYGAISG FGLSLIKWILIVRFSTYFPGYFDGQYWLWWVFLVLGFLLFLRGFINYAKVRKM PETFSNLPRTRVLFI (SEQ ID NO:242); and/or AGRYGAISGFGLSLIKWILIVRFS (SEQ ID NO:243). Also preferred are polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 5, and therefore can be used in linkage analysis as a marker for chromosome 5.

This gene is expressed primarily in kidney, adrenal gland, brain and to a lesser extent in wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, fertilization control or tissue damages by metabolites or other toxic agents. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and urosecretion system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, adrenal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zona pellucida protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for fertility control such as controceptive development. The homology with metal homeostasis and commissureless genes indicates the gene's function in spermatozoa guidance and protection. It would also be useful for the treatment/diagnosis of tissue damages caused by toxic metabolites and other agents since the gene product is also expressed in urosecretive tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in liver and to a lesser extent in placenta. Preferred polypeptide fragments comprise the amino acid sequence: MKHLSAWNFT KLTFLQLWEI FEGSVENCQTLTSYSKLQIKYTFSRGSTFYI (SEQ ID NO:244). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, digestive and nutrient transport/utilization disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and

circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., liver, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in liver and placenta indicates that the protein product is either an extracellular enzyme or a molecule carrier. Therefore, polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/treatment of digestive and nutrient transport/utilization disorders, including malabsorption and malnutrition.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene shares homology with the sap47 gene of *Drosophila melanogaster*, a gene which codes for a conserved neuronal protein associated with synaptic terminals. (See Mol. Brain Res. 32:45-54 (1995); see also, Accession No. 929571.) Thus, based on homology, the gene of the present invention also should be associated with synaptic terminals. Preferred polypeptide fragments comprise the amino acid sequence:
FSSDFRTSPWESRRVESKATSARCGLWGSGPRRRPASGMFRGLSSWLGLQQP
VAGGGQPNGDAPPEQPSETVAESAEEELQQAGDQELLHQAKDFGNYLFNFASA
ATKKITESVAETAQTIKKSVEEGKIDGIIDKTIIGDFQKEQKKFVEEQHTKKSEA
AVPPWVDTNDEETIQQQILALSADKRNFLRDPPAGVQFNFDFDQMYPVALVML
(SEQ ID NO:245); MRFALVPKLVKEEVFWRNYFYRVSLIKQSAQLTALAAQQQA
AGKGGEEQ (SEQ ID NO:246); STSPGVSEFVSDAFDACNLNQEDLRKEMEQL
VLDKKQEETAVLEEDSADWEKELQQELQEYEVVTESEKRDENWDK (SEQ ID NO:247); SPWESRRVESKATSARCGLWGSGPRRRPASGMFRGLSSWLGLQQ
PVAGGGQPNGDAPPEQPS (SEQ ID NO:248); PVAGGGQPNGDAPPEQPSETV
ESAEEELQQAGDQELLHQAKDFGNYLFNFASAATKKITESVAE (SEQ ID NO: 249); and/or FQKEQKKFVEEQHTKKSEAAVPPWVDTNDEETIQQQILALSADKR
NFLRDPPAGVQFNFDFDQMYPVALVML (SEQ ID NO:250). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in kidney pyramids and to a lesser extent in lung and other tissues of various types. This gene fluxes calcium in human aortic smooth muscle cells, and therefore is involved in signal transduction.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, renal and nervous disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and/or nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, lung, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in kidney and lung and homology with sap47 indicates that the protein product has regulatory or direct functions in molecular exchange with body fluids and nervous system signaling. Polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in kidney and nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

The translation product of this gene shares sequence homology with the mouse Ly-9.2 antigen which is thought to be an important cell surface marker in lymphoids, myeloids and hematopoietic progenitors. (See Accession No. gil198932.) Preferred polypeptide fragments comprise the amino acid sequence: PFICVARNPVSRNFSSPI LARKLCEGAA (SEQ ID NO:251); and/or KEDPANTVYSTVEIPKKMENPHSLLT MPDTPRL (SEQ ID NO:252). Also preferred are polynucleotide fragments encoding these polypeptide fragments. Based on homology, it is likely that this gene is also a cell surface marker, involved in hematopoiesis.

This gene is expressed primarily in activated macrophages, monocytes and T-cells and to a lesser extent in spleen and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., blood cells, and bone marrow, and cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:129 as residues: Lys-26 to Tyr-33, Arg-44 to Ile-49, Ser-53 to Lys-71, Lys-86 to Pro-91.

The tissue distribution and homology to Ly-9.2 surface immunoglobulin family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune and hematopoietic disorders. Polypeptides and polynucleotides corresponding to this gene are also be used as a marker for leukemia or a modulator of the functions of the cells of macrophage/monocyte or T-cell types.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene shares sequence homology with the *Drosophila* glutactin gene which is thought to be important in cell-cell interaction or cell-extracellular matrix contact.

This gene is expressed primarily in colon tissue, aorta endothelial cells and to a lesser extent in skin, breast tissue and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of these tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the gastrointestinal tract, vascular system or T-cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, cardiovascular system, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., colon, cardiovascular tissue, skin, mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to glutactin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the development and maintenance of the integrity of the basal membrane in the gastrointestinal tract and

cardiovascular system. The expression in T-cells also indicate the protein may be involved in T-cell adhesion, cell-cell interaction and development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The translation product of this gene shares sequence homology with MURF4 protein, an ATPase homolog, which is thought to be important in ATP hydrolysis.

This gene is expressed primarily in breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer and non-neoplastic breast diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to MURF4 gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neoplastic or non-neoplastic breast diseases because ATPase like protein may be involved in changed metabolic states of the breast.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

This gene shares homology to the alcohol dehydrogenase gene. Preferred polypeptide fragments comprise comprise the amino acid sequence: ASAVLLDLPNSGGEAQAKKLGNNCVFAPADVTSEKDVQTALALAKGKFGRVDVAVNCAGIAVASKTYNLKKGQTHTLEDFQRVLDVNLMGTFNVIRLVAGEMGQNEPDQGGQRGVIINTASVAAFEGQVGQAAYSASKGGIVGMTLPIARDLAPIGIRVMTIAPGLFGTPLLTSLPEKVCNFLASQVPFPSRLGDPAEYAHLVQAIIENPFLNGEVIRLDGAIRMQP (SEQ ID NO:253); and/or SVAAFEGQVGQAAYSASKGGIVGMTLPIA (SEQ ID NO:254). Polynucleotides encoding these fragements are also encompassed by the invention. Other groups have also recently cloned this gene, recognizing its homology to alcohol dehydrogenase. (See Accession No. 1778355.) Moreover, a second group

recently cloned the mouse homologue of this gene. (See Accession No. 2078284.) They found that the mouse homologue binds to amyloid beta-peptide and mediates neurotoxicity in Alzheimer's disease, calling the protein ERAB. This gene maps to chromosome X, and therefore can be used in linkage analysis as a marker for chromosome X. Therefore, mutations in the translated product of this gene may be involved in Alzheimer's disease in humans, as well as other sex linked diseases. This gene can be used as a diagnostic marker for these diseases.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:132 as residues: Arg-45 to Ser-53.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

The translation product of this gene shares week sequence homology with rat N-methyl-D-aspartate receptor subunit and other proline-rich proteins which are thought to be important in neurotransmission or protein-protein intereaction.

This gene is expressed primarily in synovial hypoxia and to a lesser extent in ovary, senescent cells and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, synovial hypoxia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the synovia and brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., synovial tissue, ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in synovial hypoxia and nerve tissues, and homology to N-methyl-D-aspartate receptor subunit and other proline-rich proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of synovial hypoxia and other synovial disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

This gene is expressed primarily in prostate and to a lesser extent in placenta and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male and female infertility, cancer, and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., prostate, placenta, ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:134 as residues: Pro-17 to Met-23, Ala-30 to Trp-38, Ile-49 to Trp-54, Lys-68 to Gly-74, Thr-93 to Gly-99, Met-126 to Glu-132, Gly-173 to Ser-178, Lys-205 to Tyr-214.

The tissue distribution of this gene in the prostate, placenta and ovary indicates that this gene product is useful for treatment/diagnosis of male or female infertility, endocrine disorders, fetal deficiencies, ovarian failure, amenorrhea, ovarian cancer, benign prostate hyperplasia, prostate cancer, and other forms of cancer of the reproductive system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in the thyroid and to a lesser extent in the pineal gland. This gene maps to chromosome 10, and therefore can be used as a marker in linkage analysis for chromosome 10. Preferred polypeptide fragments comprise the amino acid sequence: HPIEWAINAATLSQFY (SEQ ID NO:256); CWIKYCLTLMQN AQLSMQDNIG (SEQ ID NO:257); KVSYLRPLDFEEARELFLLGQHYVF (SEQ ID NO:258); MERRCKMHKRXIAMLEPLTVDLNPQ (SEQ ID NO:259); and/or SHIV KKINNLNKSALKY YQLFLD (SEQ ID NO:260). Also preferred are polynucleotides encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, thyroid and pineal gland disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., thyroid and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:135 as residues: Ser-2 to Ser-8, Thr-38 to Arg-44.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/detecting immune disorders such as arthritis, asthma, immune deficiency diseases (e.g., AIDS), and leukemia, as well as treating/detecting thymus disorders (e.g., Graves Disease, lymphocytic thyroiditis, hyperthyroidism, and hypothyroidism), and treating/detecting pineal gland disorders (e.g., circadian rhythm disturbances associated with shift work, jet lag, blindness, insomnia and old age).

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in lung and tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary or immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pulmonary tissue, and tonsils, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:136 as residues: Glu-28 to Gly-49.

The tissue distribution of this gene only in lung indicates that it could play a role in the treatment/detection of lung lymphoma or sarcoma formation, pulmonary edema and embolism, bronchitis and cystic fibrosis. Its expression in tonsils indicates a potential role in the treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases (e.g., AIDS), and leukemia, in addition to the treatment/detection of tonsillitis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in lymphoid, myeloid and erythroid cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, myeloid cells, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution of this gene in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma, immunodeficiency diseases and leukemia. Preferred embodiments of the present invention are polypeptide fragments comprising the amino acid sequence: FTHLSTCLLSLLLVRMSGFLLLARASPSI CALDSSCFVEYCSSYSSSCFLHQHFPSLLDHLCQ (SEQ ID NO:261); or FLLL ARASPSICALDSSCFVQEY (SEQ ID NO:262). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is homologous to the *Drosophila Regena* (*Rga*) gene. (See Accession No. 1658504.) This *Drosophila* gene is thought to be a homolog of the global negative

transcriptional regulator NOT2 (CDC36) from yeast, which modifies gene expression and suppresses position effect variegation. Preferred polypeptide fragments comprise the amino acid sequence: PDGRVTNIPQGMVTDQFGMIGLLTFIRAAETDPGMVHL ALGSDLTTLGLNLNS (SEQ ID NO:263); VHLALGSDLTTLGLNLNSPENLYP (SEQ ID NO:265); EDLLFYLYYMNGGDVLQLLAAVELFNRDWRYHKEERVWI TR (SEQ ID NO:264); and/or HNEDFPALPGS (SEQ ID NO:266).

This gene is expressed primarily in placenta and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurological system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:138 as residues: Leu-9 to Tyr-15, Asp-34 to Gln-46, Pro-51 to Asp-57, Gly-88 to Thr-104, Thr-123 to Ser-128.

The tissue distribution of this gene indicates that it could be used in the detection and/or treatment of neurological disorders such as such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

This gene is expressed primarily in adrenal gland tumor and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and bone disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system and in bone, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adrenal gland, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:139 as residues: Ile-52 to Trp-57.

The tissue distribution of this gene indicates that it may be involved in the treatment and/or detection of adrenal gland tumors, osteosarcomas, endocrine disorders and bone disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with the FK506 binding protein, a protein which plays an important role in immunosupression. (See Accession No. M75099.) Specifically, a 12-kDa FK506-binding protein (FKBP-12) is a cytosolic receptor for the immunosuppressants FK506 and rapamycin. (See, Proc. Natl. Acad. Sci. 88: 6677-6681 (1991).) Thus, based on homology, it is likely that this gene also has immunosuppression activity. Preferred polypeptides comprise the amino acid sequence: GRIIDTSLTRDPLVIELGQKQVIPGLEQSLLDMCVGEKRRAIIPSH LAYGKRGFPPSVPADAVVQYDVELIALIR (SEQ ID NO:267); and/or IHYTGSLV DGR IIDTS (SEQ ID NO:268). Also preferred are the polynucleotide fragments encoding these polypeptides.

This gene is expressed primarily in melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:140 as residues: Ala-118 to Phe-124, Arg-178 to Lys-201.

The tissue distribution and homology to the FK506 binding proteins which are believed to a role in immunosupression mediated by the immunosupressant drugs rapamycin and cyclosporin, indicates that this gene could serve as a novel target for the identification of novel immunosupressant drugs.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this gene shares sequence homology with the rat calcium-activated potassium channel rSK3, which is thought to be important in regulating vascular tone. (See Accession No. gi|2564072, gi|1575663, and gi|1575661.) Although homologous to these proteins, this gene contains an 18 amino acid insert, not previously identified in the homologs. Preferred polypeptide fragments comprise the amino acid sequence: CESPESPAQPSGSSLPAWYH (SEQ ID NO:269). Also preferred are the polynucleotide fragments encoding these polypeptides.

This gene is expressed primarily in B-cells, frontal cortex and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular (hyper/hypotension, asthma, pulmonary edema, pneumonia, heart disease, restenosis, atherosclerosis, stoke, angina and thrombosis) and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, and endothelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:141 as residues: Glu-72 to Gly-82, His-90 to Val-95, Gln-168 to Lys-174, Val-202 to Ser-212.

The tissue distribution and homology to calcium-activated potassium channels indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of vascular disorders (hyper/hypotension, athesma, pulmonary edema, pneumonia, heart disease, restenosis, atherosclerosis, stoke, angina and thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in smooth muscle and to a lesser extent in brain (amygdala, corpus colosum, hippocampus).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular (hypertension, heart disease, athesma, pulmonary edema, restenosis, atherosclerosis, stoke, angina, thrombosis, and wound healing), and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., smooth muscle, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:142 as residues: Lys-43 to Arg-49, Tyr-58 to Glu-65.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cadiovascular disorders (hypertension, heart disease, athesma, pulmonary edema, restenosis, atherosclerosis, stoke, angina, thrombosis, and wound healing). Expression in brain indicates a role in the treatment and diagnosis of behavioral or neurological disorders, such as depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia. and addictive behavior.

FEATURES OF PROTEIN ENCODED BY GENE NO: 19

This gene is expressed primarily in T-cells (Jurkats, resting, activated, and

anergic T-cells), endothelial cells, pineal gland, and to a lesser extent in a variety of other tissues and cell types. Preferred polypeptide fragments comprise the amino acid sequence: EEAGAGRRCSHGGARPAGLGNEGLGLGGDPDHTDTGSRSKQRINN WKESKHKVIMASASARGNQDKDAHFPPPSKQSLLFCPKSKLHIHRAEISK (SEQ ID NO:270); and/or SKQRINNWKESKHKVIMASASAR (SEQ ID NO:271). Also preferred are the polynucleotide fragments encoding these polypepides.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, immune and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neurological and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, endothelial cells, and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:143 as residues: Phe-71 to Arg-76, Pro-82 to His-87, Glu-103 to Ala-111.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immune, immuno-supressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. In addition, expression in the pineal gland might suggest a role in the diagnosis of specific brain tumors and treatment of neurological disorders. Endothelial cell expression might suggest a role in cadiovascular or respiratory/pulmonary disorders or infections (athesma, pulmonary edema, pneumonia).

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in brain and embryo and to a lesser extent in leukocytes. This gene maps to chromosome 15, and therefore can be used as a marker in linkage analysis to chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:144 as residues: Met-1 to Gly-8.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-supressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain -- and in particular the fetal brain -- would suggest a possible role in the treatment and diagnosis of developmental and neurodegenerative diseases of the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior).

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in brain, kidney, lung, liver, spleen, and a variety of leukocytes (especially T-cells) and to a lesser extent in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemias, lymphomas, autoimmune, immunosuppressive, and immunodeficiencies, hematopoietic disorders, as well as renal disorders, and neoplasms. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal, pulmonary, immune, and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

brain and other tissue of the nervous system, kidney, pulmonary tissue, liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of renal conditions, such as acture renal failure, kidney fibrosis, and kidney tubule regeneration. The expression in leukocytes and other immune tissues indicates a role in immune disorders including: leukemias, lymphomas, auto-immune, immuno-supressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain -- and in particular the fetal brain -- indicates a possible role in the treatment and diagnosis of developmental and neurodegenerative diseases of the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior).

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

This gene is expressed primarily in skin (fetal epithelium, keratinocytes and skin). This gene also maps to chromosome 19, and therefore can be used in linkage analysis as a marker for chromosome 19.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, skin cancers (e.g., melanomas), eczema, psoriasis or other disorders of the skin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., skin and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:146 as residues: Pro-28 to Glu-35, Ser-39 to Phe-44, Ala-94 to Gln-99.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of skin cancers (e.g., melanomas), eczema, psoriasis or other disorders of the skin.

FEATURES OF PROTEIN ENCODED BY GENE NO: 23

This gene maps to chromosome 11. Another group recently isolated this same gene, associating the sequence to the region thought to harbor the gene involved in Multiple Endocrine Neoplasia Type 1, or MEN 1. (See Accession No. 2529721 and Genome Res. 7(7), 725-735 (1997), incorporated herein by reference in its entirety.) Preferred polypeptide fragments comprise the amino acid sequence: LFHWACLNERA AQLPRNTAXAGYQCPSCNGPS (SEQ ID NO:272).

This gene is expressed primarily in epididymus, pineal gland, T-cells, as well as fetal epithelium, lung and kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, metabolic mediated disorders, and MEN. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, renal, neurological and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., epididymus and other reproductive tissue, pineal gland, T-cells and other blood cells, epithelium, lung, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of developmental deficiencies or abnormalities as well as a host of different disorders which arise as a result of conditions in the indicated tissues or cell types. An area of particular interest is in the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-supressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. The expression in the brain, and in particular the fetal brain, would suggest a possible role in the treatment and diagnosis of

developmental and neurodegenerative diseases of the brain and nervous system (depression, schizophrenia, Alzheimer's disease, mania, dementia, paranoia, and addictive behavior). Respiratory/pulmonary disorders, such as athesma, pulmonary edema are also potential therapeutic areas, as well as renal conditions such as acute renal failure, kidney fibrosis and kidney tubule regeneration. Moreover, this gene can be used in the treatment and/or detection of MEN I.

FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemia, lymphoma, AIDS, hematopoeitic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., spleen and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune, immuno-supressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

A closely related homolog of this gene was recently cloned by another group, calling the gene CDO, an oncogene-, serum-, and anchorage-regulated member of the Ig/fibronectin type III repeat family. (See Accession No. 2406628, and J. Cell Biol. 138(1): 203-213 (1997), herein incorporated by reference in its entirety.) Preferred polypeptide fragments comprise the amino acid sequence: FYIYYRPTDSDNDSDYKK DMVEGDKYWHSISHLQPETSYDIKMQCFNEGGESEFSNVMICETKARKSSGQP GRLPPPTLAPPQPPLPETIERPVGTGAMVARSSDLPYLIVGVVLGSIVLIIVTFIPF CLWRAWSKQKHTTDLGFPRSALPPSCPYTMVPLGGLPGHQAVDSPTSVASVD

GPVLM (SEQ ID NO:273); or YIYYRPTDSDNDSDYKKDMVEGDKYWHSISHLQ PETSYDIKMQCFNEGGESEFSNVMICETKARKS (SEQ ID NO:274).

This gene is expressed primarily in fetal lung and kidney, human embryo and osteoclastoma stromal cells and to a lesser extent in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders and cancers, as well as pulmonary and renal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory/pulmonary, skeletal and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, kidney, embryonic tissue, and bone cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:149 as residues: Thr-5 to Pro-18, Ala-76 to Thr-84.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of: osteoperosis, fracture, osteosarcoma, ossification, and osteonecrosis, as well as respiratory/pulmonary disorders, such as athesma, pulmonary edema, and renal conditions such as acute renal failure, kidney fibrosis and kidney tubule regeneration.

FEATURES OF PROTEIN ENCODED BY GENE NO: 26

This gene is homologous to the HIV envelope glycoprotein. (See Accession No. 2641463.) Preferred polypeptide fragments comprise the amino acid sequence: NVRALLHRMPEPPKINTAKFNNNKRKNLSL (SEQ ID NO:275).

This gene is expressed primarily in pineal gland and skin, and to a lesser extent in lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, neurological and behavior disorders; respiratory/pulmonary disorders, such as athesma, pulmonary edema; skin conditions such as eczema, psoriasis, acne and skin cancer, as well as AIDS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and respiratory systems, as well as skin and AIDS, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, pineal gland, epidermis, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:150 as residues: Gln-15 to Gln-20.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions which affect the above tissues, such as: skin cancer, eczema, psoriasis, acne, athesma, pulmonary edema, neuro-degenerative or developmental disorders such as Alzheimer's, depression, schizophrenia, dementia, and AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

Preferred polypeptide encoded by this gene comprise the following amino acid sequence: NTNQREALQYAKNFQPFALNHQKDIQVLMGSLVYLRQGIENSPYVHL LDANQWADICDIFTRDACALLGLSVESPLSVSFSAGCVALPALINIKAVIEQRQC TGVWNQKDELPIEVDLGKKCWYHSIFACPILRQQTTDNNPPMKLVCGHIISRD ALNKMFNGSKLKCPYCPMEQSPGDAKQIFF (SEQ ID NO:276). Polynucleotides encoding such polypeptides are also provided as are complementary polynucleotides thereto.

This gene is expressed primarily in liver (adult and fetal) and spleen tissue, and to a lesser extent in placenta, T helper cells, kidney tumor, ovarian tumor, melanocytes and fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and developmental diseases and disorders and liver diseases such as liver cancer. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, circulatory and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, placenta, blood cells, kidney, ovary and other reproductive tissue, melanocytes, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of growth, hematopoietic and immune system disorders particularly related to the liver.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

The translation product of this gene shares sequence homology with prostaglandin transporter which is thought to be important in metabolic and endocrine disorders. See, for example, Gastroenterology Oct:109(4):1274-1282 (1995). Preferred polypeptides encoded by this gene comprise the following amino acid sequence: SYLSACFAGCNSTNLTGCACLTTVPAENATVVPGKCPSPGCQEAFLTFLCVMCI CSLIGAMARHP (SEQ ID NO:277); and/or PSVIILIRTVSPELKSYALGVLFLLLRL LGFIPPPLIFGAGIDSTCLFWSTFCGEQGACVLYDNVVYRYLYVSIAIALKSFAFI (SEQ ID NO:278).

This gene is expressed primarily in hematopoietic and brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic, immune and endocrine diseases and disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., endocrine tissue, hematopoietic tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to prostaglandin (and anion) transporter indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of endocrine, metabolic, immune and kidney disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed primarily in early stage human lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developmental and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:153 as residues: Val-50 to Trp-55.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of respiratory and growth diseases and disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 30

The translation product of this gene shares sequence homology with human DNA helicase which is thought to be important in accurate and complete DNA replication in creation of new cells. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: QSLFTRFVRVGVPTVDLDAQGRARA SLCXXYNWRYKNLGNLPHVQLLPEFSTANAGLLYDFQLINVEDFQGVGESEPN PYFYQNLGEAEYVVALFMYMCLLGYPADKISILTTYNGQKHLIRDIINRRCGNN PLIGRPNKVTTVDRFQGQQNDYILLSLVRTRAVGHLRDVRRLVVAMSRAR (SEQ ID NO:279); and/or LVKEAKIIAMTCTHAALKRHDLVKLGFKYDNILMEE AAQILEIETFIPLLLQNPQDGFSRLKRWIMIGDHHQLPPVI (SEQ ID NO:280).

This gene is expressed primarily in testes tumor and to a lesser extent in adrenal

gland tumor and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers and endocrine/growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine, developmental, and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., testes and other reproductive tissue, adrenal gland, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to DNA helicase indicates that the protein products of this gene are useful for study, treatment, and diagnosis of many cancer types, including testicular cancer; as well as disorders involving endocrine function and normal growth and development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The translation product of this gene shares sequence homology with BID-apoptotic death gene (mouse), Genbank accession no. PID g1669514, which is thought to be important in programmed cell death.

This gene is expressed primarily in jurkat membrane bound polysomes and activated neutrophils and to a lesser extent in endothelial cells and human cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers and other proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, endothelium, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:155 as residues: Glu-4 to Leu-11, Cys-28 to Arg-35, Gln-50 to His-66, Glu-73 to Gln-79, Gly-94 to Ser-100, Arg-114 to Asp-126, Pro-139 to Lys-146.

The tissue distribution and homology to BID-apoptotic death gene indicates that the protein products of this gene are useful for study of cell death, and treatment and diagnosis of proliferative disorders and cancers. Apoptosis - programmed cell death - is a physiological mechanism involved in the deletion of peripheral T lymphocytes of the immune system, and its dysregulation can lead to a number of different pathogenic processes. Diseases associated with increased cell survival, or the inhibition of apoptosis, include cancers (such as follicular lymphomas, carcinomas with p53 mutations, and hormone-dependent tumors, such as breast cancer, prostrate cancer, Kaposiís sarcoma and ovarian cancer); autoimmune disorders (such as systemic lupus erythematosus and immune-related glomerulonephritis rheumatoid arthritis) and viral infections (such as herpes viruses, pox viruses and adenoviruses), inflammation; graft vs. host disease, acute graft rejection, and chronic graft rejection. Diseases associated with increased apoptosis include AIDS; neurodegenerative disorders (such as Alzheimer's disease, Parkinson's disease, Amyotrophic lateral sclerosis, Retinitis pigmentosa, Cerebellar degeneration); myelodysplastic syndromes (such as aplastic anemia), ischemic injury (such as that caused by myocardial infarction, stroke and reperfusion injury), toxin-induced liver disease (such as that caused by alcohol), septic shock, cachexia and anorexia. Thus, the invention provides a method of enhancing apoptosis in an individual by treating the individual with a polypeptide encoded by this gene.

FEATURES OF PROTEIN ENCODED BY GENE NO: 32

The translation product of this gene shares sequence homology with human fructose transporter which is thought to be important in normal metabolic function and activity.

This gene is expressed primarily in T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, leukemia and other cancers, and metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic, lymph and metabolic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:156 as residues: Pro-22 to Gly-48, Ser-54 to Pro-61.

The tissue distribution indicates that the protein products of this gene are useful for study of mechanisms leading to cancer, treatment and diagnosis of cancerous and pre-cancerous conditions; as well as the study and treatment of various metabolic diseases and disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is expressed primarily in human meningima.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and other disorders of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., meningima and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:157 as residues: Asn-23 to Pro-31.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of disorders of the CNS and inflammatory responses.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene is expressed primarily in activated monocytes and wound healing tissues and to a lesser extent in fetal epithelium.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and inflammatory disorders and wound healing and tissue repair dysfunctions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, epithelial and gastrointestinal systems, and healing wounds, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., monocytes and other blood cells, and epithelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:158 as residues: Ala-28 to Ala-33, Gly-35 to Glu-45.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis, study and treatment of immune and inflammatory disorders and wound healing dysfunctions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

This gene is expressed primarily in human osteosarcoma and prostate cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, skeletal and neoplastic conditions such as bone and prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bone, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:159 as residues: Ser-14 to Gly-22, Leu-37 to Gln-43.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of skeletal disorders and cancer..

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene encodes a protein which is highly homologous to a protein called congenital heart disease protein 5, presumably implicated in congenital heart disease (see Genbank PID g2810996).

This gene is expressed primarily in Hodgkin's lymphoma, erythroleukemia cells, and TNF activated synovial fibroblasts, to a lesser extent in ovarian cancer, cerebellum, spleen, fetal liver and placenta and finally to a lesser extent in various other mesenchymal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, immune, hematopoietic and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, hematopoietic and cardiovascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., heart and other cardiovascular tissue, lymphoid tissue, blood cells, bone marrow, ovary and other reproductive tissue, brain and other tissue of the nervous system, spleen, liver, and mesenchymal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:160 as residues: Lys-41 to Met-49, Gln-54 to Glu-59, Glu-76 to Thr-88.

The homology of this gene and translation product to congenital heart disease protein 5 indicates a role for this protein in the diagnosis, prognosis and/or treatment of

heart disease or other cardiovascular related disorders. In addition, predominant expression in cells associated with the immune and hematopoetic system indicates a role for this protein in the treatment, diagnosis and/or prognosis of immune and autoimmune diseases, such as lupus, transplant rejection, allergic reactions, arthritis, asthma, immunodeficiency diseases, leukemia, AIDS, thymus disorders such as Graves Disease, lymphocytic thyroiditis, hyperthyroidism and hypothyroidism, graft versus host reaction, graft versus host disease, transplant rejection, myelogenous leukemia, bone marrow fibrosis, and myeloproliferative disease. The protein could also be used to enhance or protect proliferation, differentiation and functional activation of hematopoietic progenitor cells such as bone marrow cells, which could be useful for cancer patients undergoing chemotherapy or patients undergoing bone marrow transplantation. The protein may also be useful to increase the proliferation of peripheral blood leukocytes, which could be useful in the combat of a range of hematopoietic disorders including immunodeficiency diseases, leukemia, and septicemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, urogenital neoplasias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:161 as residues: Asn-22 to Asn-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of ovarian and other tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

The translation product of this gene shares sequence homology with zinc finger proteins.

This gene is expressed primarily in various fetal, cancer, and endothelial lines.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of immune and developmental conditions and cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

This gene is expressed primarily in fetal, infant, and adult brain and to a lesser extent in other brain and endocrine organs and blastomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain tumors and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, endocrine tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

disorder.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of brain cancer and other neurological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

The translation product of this gene shares sequence homology with vesicular glycoproteins and lectins. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: DTYPNEEKQQERVFPXXSAMVNNGSLSYDHER DGRPTELGGCXAIVRNLHYDTFLVIRYVKRHLTIMMDIDGKHEWRDCIEVPGV RLPRGYYFGTSSITGDLSDNHDVISLKLFELTVERTPEEE (SEQ ID NO:281); and/or LKREHSLSKPYQGVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQGA LWNRVPCFLRDWELQVHFKIHGQGKKNLHGDGLAIWYT (SEQ ID NO:282).

This gene is expressed primarily in infant brain and to a lesser extent in various normal and transformed neural, endocrine, and immune organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and neurodevelopmental conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hormonal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, endocrine tissue, and tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:164 as residues: Pro-64 to Gly-71, Gly-94 to Leu-100, Thr-110 to Pro-116, Thr-135 to Arg-145, Glu-164 to Glu-171, Asp-204 to Asp-211, Arg-253 to His-261, Asn-312 to Tyr-323.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of mental retardation and other neurological disorders and neoplasias.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene displays homology to the glycosyltransferase family, which catalyze the addition of sialic acids to carbohydrate groups which are present on glycoproteins.

This gene is expressed primarily in smooth muscle and to a lesser extent in pineal gland, fetal liver, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, gastrointestinal injury, inflammatory and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., smooth muscle, pineal gland, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:165 as residues: Ser-12 to Trp-21, Arg-24 to Pro-32, Asp-73 to Lys-82, Lys-90 to Ala-97.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of neurodegenerative and growth disorders and gastrointestinal repair.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

The translation product of this gene shares sequence similarity with metallothionein polypeptides. See, for example, Proc. Natl. Acad. Sci. U S A 1992 Jul 15:89(14):6333-6337. Metallothioneins are believed to inhibit neuronal survival among other biological functions. Based on the sequence similarity (especially the conserved cysteine motifs characteristic of the metallothionein family) the translation product of this gene is expected to share certain biological activities with other members of the metallothionein polypeptide family. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: PGTLQCSALHHDPGCANCSRFCRD CSPPACQC (SEQ ID NO:283).

This gene is expressed exclusively in placenta and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, liver, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to metallothionien indicates that the protein products of this gene are useful for diagnosis and treatment of immune and hematopoietic system disorders and neurological diseases, especially in fetal development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 43

Preferred polypeptides encoded by this gene comprise the following amino acid sequence: FLYDVLMXHEAVMRTHQIQLPDPEFPS (SEQ ID NO:284).

This gene is expressed primarily in T-cells and synovial tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment and diagnosis of disorders of the immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 44

The translation product of this gene shares sequence similarity with several methyltransferases (e.g., see Genbank gi|1065505).

This gene is expressed primarily in ovary, thymus, infant adrenal gland, tissues of the nervous system and the hematopoietic tissue, and to a lesser extent in adipose tissue and many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the reproductive system, the endocrine system, the hematopoietic system and the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, endocrine, CNS and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., ovary and other reproductive tissue, thymus, adrenal gland, brain and other tissue of the nervous system, hematopoietic tissue, and adipose tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:168 as residues: Ser-3 to Gly-12, Asp-19 to Arg-31, Tyr-70 to Tyr-77, Asn-130 to Lys-140, Pro-165 to Gln-170, Pro-192 to Lys-199, Leu-216 to Glu-227, Glu-254 to Phe-281.

The tissue distribution and homology to methyltransferase indicates that the protein products of this gene are useful for diagnosis and treatment of disorders of the CNS, the hematopoietic system and reproductive organs and tissues. For example, the abundant expression in the ovary may indicate that the gene product can be used as a hormone with either systemic or reproductive functions; as growth factors for germ cell maintenance and in vitro culture; as a fertility control agent; remedy for sexual dysfunction or sex development disorders: diagnostics/treatment for ovarian tumors, such as serous adenocarcinoma, dysgerminoma, embryonal carcinoma,

choriocarcinoma, teratoma, etc; The expression in thymus may indicate its utilities in T-cell development and thus its applications in immune related medical conditions, such as infection, allergy, immune deficiency, tissue/organ transplantation, etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with cytochrome C oxidase which is thought to be important in metabolic function of cells. This gene has now recently been published as estrogen response gene. See Genbank accession no. AB007618 and Mol. Cell. Biol. 18 (1), 442-449 (1998). See also J Immunol. Mar 1:154(5): 2384-2392 (1995), where the mouse homologue was published and implicated in siliocis.

This gene is expressed primarily in adipose tissue, kidney and fetal brain and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic diseases involving especially adipose tissue, brain and kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS and vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adipose tissue, kidney, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:169 as residues: Thr-5 to Ser-14.

The tissue distribution and homology to cytochrome C oxidase, estrogen response gene product and siliocis related gene product indicates that the protein products of this gene are useful for diagnosis and treatment of metabolic disorders in the CNS, adipose tissue and kidney, particularly siliocis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 46

The translation product of this gene shares sequence homology with reticulocalbin. See, for example, J. Biochem. 117 (5), 1113-1119 (1995). Based on the

sequence similarity, the translation product of this gene is expected to share certain biological activities with reticulocalbin, e.g., Ca++ binding activities. This gene product is sometimes hereinafter referred to as "Reticulocalbin-2".

This gene is expressed primarily in breast, endothelial cells, synovial, heart and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the breast, vascular and skeletal/cardiac muscular system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast, vascular and skeleto-muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, endothelial cells, synovial tissue, heart and other cardiovascular tissue, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:170 as residues: Gly-16 to Arg-32, Ala-42 to Asn-50, Glu-66 to Gln-76, Arg-85 to Gly-94, Thr-108 to Asp-115, Trp-121 to Gly-130, Leu-137 to His-144, Glu-155 to Lys-161, Asp-175 to Ser-180, Glu-209 to Gly-217, Glu-232 to Glu-237, Thr-243 to Asp-261, Glu-287 to Arg-295.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of diseases of the vascular and skeletal/cardiac muscular system. The homology of the gene with reticulocalbin indicates its biological function in regulating calcium store, a particularly important function in muscular cell types. The gene expression in the heart may indicate its utilities in diagnosis and remedy in heart failure, ischemic heart diseases, cardiomyopathy, hypertension, arrhythmia, etc. The abundant expression in the breast may indicate its applications in breast neoplasia and breast cancers, such as fibroadenoma, papillary carcinoma, ductal carcinoma, Pagetís disease, medullary carcinoma, mucinous carcinoma, tubular carcinoma, secretory carcinoma and apocrine carcinoma; juvenile hypertrophy and gynecomastia, mastitis and abscess, duct ectasia, fat necrosis and fibrocystic diseases. etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene shares weak sequence homology with H+-transporting ATP synthase which is thought to be important in cell metabolism or signal transduction.

This gene is expressed only in testis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of some types of diseases and conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Since only one out of about a million expressed sequence tag is found in testes indicates that its expression is low and selectively in testes. Since some of the genes only expressed in testes are usually expressed in brain or in certain induced hematopoietic cells/tissues, it is speculated that this gene to be expressed in brain or hematopoietic cells/tissues and is useful for diagnosis and treatment of disorders these systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with human polymeric immunoglobulin receptor (accession No.X73079) which is thought to be important in antibody recognition and immune defenses. In one embodiment, polypeptides of the invention comprise the sequence GWYWCG (SEQ ID NO:285). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta and to a lesser extent in corpus callosum and fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, disorders of the immune system, e.g. autoimmune diseases and immunodeficiency. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:172 as residues: Tyr-37 to Cys-49, Gly-51 to Tyr-56, Lys-88 to Trp-93, Leu-130 to Glu-136.

The tissue distribution and homology to human polymeric immunoglobulin receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, e.g. autoimmune diseases and immunodeficiencies.

FEATURES OF PROTEIN ENCODED BY GENE NO: 49

This gene is expressed in thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., thymus and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, e.g. autoimmunity and immunodeficiency.

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

Preferred polypeptide encoded by this gene comprise the following amino acid sequence: MKVGARIRVKMSVNKAHPVVSTHWRWPAEWPQMFLHLAQEPRTE VKSRPLGLAGFIRQDSKTRKPLEQETIMSAADTALWPYGHGNREHQENELQKY LQYKDMHLLDSGQSLGHTHTLQGSHNLTALNI (SEQ ID NO:286). Polynucleotides encoding this polypeptide are also provided as are complementary polynucleotides thereto.

This gene is expressed primarily in adrenal gland, pituitary, T helper cells, and breast cells and to a lesser extent in a wide variety of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of the some diseases and conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adrenal gland, pituitary, T-cells and other blood cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:174 as residues: Gln-39 to Ser-47, Arg-57 to Glu-67, Tyr-82 to Gln-95.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of a wide range of disorders, such as immune and endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 51

The translation product of this gene shares sequence homology with human Sop2p-like protein which is important in cytoskeleton structure. In one embodiment, polypeptides of the invention comprise the sequence SLHKNSVSQISVLSGGKAKCS QFCTTGMDGGMSIWDVKSLESALKDLKI (SEQ ID NO:287). Polynucleotides encoding this polypeptide are also encompassed by the invention. This gene maps to chromosome 7. Therefore, polynucleotides of the invention can be used in linkage

analysis as a marker for chromosome 7.

This gene is expressed primarily in immune and hematopoietic tissues/cells and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and hematopoietic disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune and hematopoietic tissue/cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:175 as residues: Lys-49 to Gln-54, Ala-61 to Arg-66, Lys-82 to Lys-87, Glu-126 to Val-133, His-136 to Ile-141, Glu-175 to Ser-187, Asp-286 to Leu-296, Ala-298 to Ser-310.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immunological, hematopoietic, and inflammatory disorders, e.g, immunodeficiency, autoimmunity, inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

The translation product of this gene shares sequence homology with *Caenorhabditis elegans* R53.5 gene encoding a putative secreted protein without known function.

This gene is expressed primarily in endothelial cells, brain and several highly vascularized, and tumor tissues and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, aberrant angiogensis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular and brain system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, brain and other tissue of the nervous system, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:176 as residues: Thr-43 to Asn-60, Thr-106 to Phe-115, Asp-122 to Arg-133, Arg-186 to Asp-192, Leu-211 to Lys-216.

The tissue distribution and homology to a *C. elegans* secreted protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of disorders in vascular or brain system, e.g. aberrant angiogenesis, ischemia, neurodegeneration, etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

In one embodiment, polypeptides of the invention comprise the sequence EASKSSHAGLDLFSVAACHRF (SEQ ID NO:288). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in T-cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphocytic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lymphoid system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:177 as residues: Pro-3 to Thr-8, Arg-37 to Asp-46.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment, and cure of lymphocytic disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

The translation product of this gene shares sequence homology with secreted cartilage matrix protein, a major component of the extracellular matrix of nonarticular cartilage which is thought to be important in cartilage structure. In specific embodiments, polypeptides of the invention comprise the sequence: RCKKCTEGPI DLVFVIDGSKSLGEENFEVVKQF (SEQ ID NO:297); VTGIIDSLTISPKAARVGL LQYSTQVH (SEQ ID NO:290); TEFTLRNFNSAKDMKKAVAHMKYM (SEQ ID NO:291); GKGSMTGLALKHMFERSFTQGEGARPF (SEQ ID NO:292); STRVP RAAIVFTDGRAQDDVSEWASKAKANGITMYAVGVGKAIE (SEQ ID NO:293); EELQEIASEPTNKHLFYAEDFSTMDEISEKLKKGICEALEDS (SEQ ID NO:294); TQRLEEMTQRM (SEQ ID NO:295); PQGCPEQPLH (SEQ ID NO:296); and/or YMGKGSMTGLALKHMFERSFT (SEQ ID NO:289). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta, infant brain, prostate, fetal lung and to a lesser extent in endometrium and fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormal placenta and pregnancy, disorder and injury in brain, prostate, and vasculature. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction, neuronal, and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., placenta, brain and other tissue of the nervous system, prostate, lung and endometrium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to cartilage matrix protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis,

treatment, and cure of abnormalities in placenta and pregnancy, disorder and injury in brain, prostate, and vasculature.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

The translation product of this gene is the human ortholog of bovine and hamster CII-3, a succinate-ubiquinone oxidoreductase complex II membrane-intrinsic subunit, which is thought to be important in mitochondrial electron transport chain during metabolism. In specific embodiments, the polypeptides of the invention compriseMAALLLRHVGRHCLRAHFSPQLCIRNAVPLGTTAKEEMERFWNKNIG SNRPLSPHITIYS (SEQ ID NO:298); VFPLMYHTWNGIRHLMWDLGKGLKIPQL YQSG (SEQ ID NO:299); MAALLLRHVGRHCLRAH (SEQ ID NO:300); VKSLCL GPALIHTAKFAL (SEQ ID NO:301); VFPLMYHTWNGIRHLMWDLGKGL (SEQ ID NO:302).

This gene is expressed in 8-week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolism disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment, and cure of metabolism disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

This gene is expressed primarily in umbilical vein endothelial cells, human ovarian tumor cells, human meningima cells, and human Jurkat membrane bound polysomes. In specific embodiments, polypeptides of the invention comprise the amino acid sequence: RVWDVRPFAPKERCVKIFQGNV (SEQ ID NO:303); HNFEKNLL

RCSWSPDGSKIAAGSADRFVYV (SEQ ID NO:304); and/or WDTTSRRILYKLPG HAGSINEVAFHPDEPI (SEQ ID NO:305). Polynucleotides encoding these polypeptides are also encompassed by the invention.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, immune and cardiovascular disorders and urogenital neoplasias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of these tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neurological, urogenital, reproductive system and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, cells, endothelial cells, ovary and other reproductive tissue, meningima, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:143 as residues: Phe-71 to Arg-76, Pro-82 to His-87, Glu-103 to Ala-111.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immune, immuno-supressive (e.g. transplantation) and immunodeficiencies (e.g. AIDS) and hematopoietic disorders. In addition, expression in ovarian tumor cells suggests that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of ovarian tumors, and other tumors and neoplasias. Further, endothelial cell expression suggests a role in cadiovascular or respiratory/pulmonary disorders or infections (athsma, pulmonary edema, pneumonia).

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

The translation product of this gene shares sequence homology with type I collagen. In specific embodiments, the polypeptides of the invention comprise the sequence: GRIPAPAPSVPAGPDSR (SEQ ID NO:309); VRGRTVLRPGLDAEPE LSPE (SEQ ID NO:306); EQRVLERKLKKERKKEERQ (SEQ ID NO:307); ARRSG

AELAWDYLCRWAQKHKNWRFQKTRQTWLLLHMYDSDKVPDEHFSTLLAYLE GLQGR (SEQ ID NO:255); and/or RLREAGLVAQHPP (SEQ ID NO:308). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in epididymus, prostate cell line (LNCAP), and pituitary gland; and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the epididymus, prostate (especially prostate cancer), and pituitary gland. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system and neuroendocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., epididymus and other reproductive tissue, prostate, and pituitary gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to type I collagen, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the epididymus, prostate (especially prostate cancer), and pituitary gland.

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

This gene is expressed primarily in the frontal cortex of the brain from a schizophrenic individual.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of

the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of schizophrenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

The polypeptide encoded by Gene 59 is homologous to human surface 4 integral membrane protein. In specific embodiments, the polypeptides of the invention comprise the sequence: TGCVLVLSRNFVQYACFGLFGIIALQTIAYSILWDLKF LMRN (SEQ ID NO:310); SRSEGKSMFAGVPTMRESSPKQYMQLGGRVLLV LMFMTLLHFDASFFSIVQNIVG (SEQ IDNO:311); GTAEDFADQFLRVTKQYLP HVARLCLISTFLEDGIRMFQWSEQRDYIDTTWNCGYLLAS (SEQ ID NO:312); LMRNESRS (SEQ ID NO:314); ASFLLSRTSWGTA (SEQ ID NO:315); and/or ASFLLSRTSWGTALMIL (SEQ ID NO:313). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Hodgkin's lymphoma and lung; and to a lesser extent in many other human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma, tumors or other abnormalities of the lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., lymphoid tissue, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:183 as residues: Met-20 to Trp-27.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of Hodgkin's lymphoma, tumors or other abnormalities of the lung.

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

This gene is expressed primarily in bone cancer and stomach cancer, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone cancer and stomach cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, and the stomach, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bone, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of bone cancer and stomach cancer and possibly other cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 61

This gene is expressed primarily in epididymus, and lymph node of breast cancer, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the epididymus, and breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the epididymus and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., epididymus and other reproductive tissue, lymphoid tissue, and mammary tissue, and cancerous and wounded tissues) or bodily fluids

(e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:185 as residues: Arg-57 to Ser-65.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the epididymus, and breast cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

The translation product of this gene appears to be the human homolog of bovine NADH dehydrogenase which is thought to be important in cellular metabolism. In specific embodiments, the polypeptides of the invention comprise the amino acid sequence: SMSALTRLASFARVGGRLFRSGCARTAGDGGVRHAGGGVHIEPRY RQFPQLTRSQVFQSEFFSGLMWFWILWRFWHDSEEVLGHFPYPDPSQWTDEEL GIPPDDED (SEQ ID NO:323), or fragments thereof. Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed in larynx tumor, lymph node, brain amygdala, human cardiomyopathy, and retina.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases affecting cellular metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., larynx, lymphoid tissue, brain and other tissue of the nervous system, heart and cardiovascular tissue, and retina, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:208 as residues: Pro-27 to Gln-32, Arg-42 to Glu-51.

The tissue distribution and homology to NADH dehydrogenase indicates that

polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases involving cellular metabolism.

FEATURES OF PROTEIN ENCODED BY GENE NO: 63

This gene is expressed primarily in amygdala, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, abnormalities of the amygdala. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., amygdala, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:187 as residues: Gln-17 to Glu-29, Pro-41 to Phe-46, Ser-59 to Ile-70, Thr-97 to Leu-105.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of amygdala.

FEATURES OF PROTEIN ENCODED BY GENE NO: 64

This gene is expressed primarily in female bladder, and to a lesser extent in chronic synovitis and hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bladder cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the urinary tract, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bladder, synovial tissue, and

vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:188 as residues: Pro-2 to Gln-7, Pro-27 to Phe-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatments of defects of the urinary tract, especially bladder cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 65

This gene is expressed primarily in fetal spleen, and to a lesser extent in hemangiopericytoma, thymus, and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, defects of immune of hematopoietic systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune of hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., spleen, vascular tissue, thymus, blood cells, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The protein product of this gene is useful for treatment of defects of the immune or hematopoietic systems, because of the gene's expression in thymus and spleen.

FEATURES OF PROTEIN ENCODED BY GENE NO: 66

This gene is expressed primarily in human pituitary and to a lesser extent in placenta and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, endocrine growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., pituitary and other endocrine tissue, placenta, and pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:190 as residues: Val-38 to Asn-44, Gly-53 to Ser-65.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of growth disorders related to pituitary dysfunction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

The translation product of this gene shares sequence homology with a *Caenorhabditis elegans* gene of unknown function. In specific embodiments, the polypeptides of the invention comprise the sequence: DPRRPNKVLRYKPPPSE CNPALDDPTP (SEQ ID NO:317); DYMNLLGMIFSMCGLMLKLKWCAWVA VYCS (SEQ ID NO:318); FISFANSRSSEDTKQMMSSF (SEQ ID NO:316); and/or MLSISAVVMSYLQNPQPMTPPW (SEQ ID NO:319). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in primary breast cancer and lymph node breast cancer and to a lesser extent in adult brain, lung cancer, colon cancer, epithelioid sarcoma, and Caco-2 cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, lymphoid tissue, brain and other tissue of the nervous system, lung, colon, and

epithelium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:191 as residues: Asn-34 to Lys-42.

The tissue distribution in a variety of cancer tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of a variety of cancer and tumor types.

FEATURES OF PROTEIN ENCODED BY GENE NO: 68

The translation product of this gene shares sequence homology with steroid membrane binding protein. The translation product of this gene has recently been published as progesterone binding protein. See Genbank AJ002030. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: AAGDGDVKLGTLGSGSESSNDGGSESPGDAGAAAXGGGWAAAALALLTG GGE (SEQ ID NO:320).

This gene is expressed primarily in breast, and to a lesser extent in placenta and fetal tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer or developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of breast or fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., mammary tissue, placenta, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:192 as residues: Pro-43 to Asp-49, Gln-54 to Pro-64, Asp-110 to Asp-118, Lys-138 to Tyr-143, Pro-150 to Asp-170.

The tissue distribution and homology to steroid membrane binding protein and to progesterone binding protein indicates that the protein products of this gene are

useful for treatment of breast cancers, especially those caused by estrogen and progesterone binding.

FEATURES OF PROTEIN ENCODED BY GENE NO: 69

Preferred polypeptides encoded by this gene comprise the following amino acid sequence: AADNYGIPRACRNSARSYGAAWLLLXPAGSSRVEPTQDISISDQLGG QDVPVFRNLSLLVVGVGAVFSLLFHLGTRERRRPHAXEPGEHTPLLAPATAQPL LLWKHWLREXAFYQVGILYMTTRLIVNLSQTYMAMYLTYSLHLPKKFIATIPLV MYLSGFLSSFLMKPINKCIGRN (SEQ ID NO:321).

This gene is expressed primarily in macrophage (GM-CSF treated), and to a lesser extent in monocytes and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and infection . Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., macrophages and other blood cells, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment of infection or inflammation or other events or defects involving the immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

This gene is expressed primarily in adult brain and to a lesser extent in thyroid, 12 week old early stage human, and stromal cell TF274.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological or neuro-endocrine diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, thyroid, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:194 as residues: Pro-65 to Cys-71.

The tissue distribution indicates that the protein products of this gene are useful for treatment and diagnosis of neurological diseases or metabolic conditions involving the neuro-endocrine system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed in T-cell helper and to a lesser extent in adult brain and adult testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, meningitis or reproductive problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, neural and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, brain and other tissue of the nervous system, testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:195 as residues: Val-18 to Tyr-24, Ala-89 to Asp-99, Asp-104 to Ala-117, Leu-121 to Pro-136.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis immune and

reproductive disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 72

The translated polypeptide of this contig has a high degree of identity with the Ob Receptor-Associated Protein deposited as GenBank Accession No. 2266638. No function has been determined for the Ob Receptor-Associated Protein, however it is expressed upon stimulation of the Ob Receptor by Leptin.

This gene is expressed in T-cells and to a lesser extent in endothelial and bone marrow cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, acute lymphoblastic leukemia, hematapoetic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematapoetic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, endothelial cells, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:196 as residues: Ser-61 to Trp-70.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of leukemia and other disorders of the primary immune system. In addition, since this gene appears to be related to the Ob Receptor-Related Protein, it is likely that this polypeptide is also involved in the Ob/Leptin signal transduction cascade. As a result, this protein may be of use in the molecular diagnosis and therapeutic intervention of obesity and related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 73

The translation product of this contig has homology with furin, a protein thought to be a key endopeptidase in the constitutive secretory pathway. The identification and initial characterization of Furin was reported by Takahasi and

colleagues (Biochem Biophys Res Commun 1993 Sep 15;195(2):1019-1026).

This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system such as allergies, wound healing and antigen recognition. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neutrophils and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of allergies or other immune disorders since neutrophils are an important part of an allergic response. Further, since this protein appears to be related to Furin, it can be used diagnostically and therapeutically to treat secretory protein processing disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed in the frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, of the motor activity and sensory functions that involve the central nervous system . Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of neural disorders that affect cognitive functions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

The translation product of this gene shares sequence homology with inorganic pyrophophatase which is thought to be important in the catalysis the hydrolysis of diphosphate bonds, chiefly in nucleoside di- and triphosphates and essential enzymes that are important for controlling the cellular levels of inorganic pyrophosphate (PPi). The bovine homolog of this gene has been identified by Yang and Wensel (J. Biol. Chem. 267:24641-24647 (1992)).

This gene is expressed in osteoclastoma cells and to a lesser extent in epithelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoporosis and other bone weakening diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, and epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:199 as residues: Lys-22 to Tyr-28, Asp-64 to Lys-77, Pro-86 to Ile-91, Gln-99 to Pro-119, Tyr-169 to Asp-174, Lys-176 to Gly-181, Trp-189 to Asn-202, Lys-233 to Gly-239, Ser-250 to Asp-257.

The tissue distribution and homology to inorganic pyrophophatase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of osteoporosis through the removal of bone by demineralization.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

The translation product of this gene shares exact sequence homology with ATP sulfurylase/APS kinase (GenBank Accession No. 2673862) which is thought to be important in biosynthesis of the activated sulfate donor, adenosine 3'-phosphate 5'-phosphosulfate, involves the sequential action of two enzyme activities: ATP sulfurylase, which catalyzes the formation of adenosine 5'-phosphosulfate (APS) from ATP and free sulfate, and APS kinase, which subsequently phosphorylates APS to produce adenosine 3'-phosphate 5'-phosphosulfate.

This gene is expressed in osteoclastoma cells and to a lesser extent in developmental tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, antibiotic resistant bacterial infections, osteoarthritis and other auto immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or skeletal structure expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., bone, and developmental tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:200 as residues: Asn-15 to Trp-20, Ser-36 to Gly-41, Pro-103 to Val-110, Pro-134 to Arg-143, Leu-173 to Arg-178, Ser-190 to Ala-197, His-314 to Arg-319, Arg-354 to Asn-362, Asp-391 to Arg-397, Glu-402 to Asp-409, Asp-434 to Leu-439, Glu-441 to Arg-446, Gly-455 to Asp-462, Pro-528 to His-541, Asn-566 to Arg-571, Tyr-574 to Glu-581, Thr-589 to Glu-603.

The tissue distribution and homology to ATP sulfurylase/APS kinase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment or detection of autoimmune diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This polypeptide is identical to the SLP-76-associated protein reported by Musci and colleagues (J. Biol. Chem. 272 (18), 11674-11677 (1997)) and to the FYB protein

reported by da Silva and coworkers (Proc. Natl. Acad. Sci. U.S.A. (1997) In press). These proteins have been reported to be novel T-cell Proteins which bind FYN and SLP-76 and regulate IL-2 production. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: RITDNPEGKWLGRTARGSYGYIK TTAVEIXYDSLKLKKDSLGAPSRPIEDDQEVYDDVAEQDDISSHSQSGSGGIFPP PPDDDIYDGIEEEDADDGFPAPPKQLDMGDEVYDDVDTSDFPVSSAEMSQGTNV GKAKTEEKDLKKLKKQXKEXKDFRKKFKYDGEIRVLYSTKVTTSITSKKWGT RDLQVKPGESLEVIQTTDDTKVLCRNEEGKYGYVLRSYLADNDGEIYDDIADGC IYDND (SEQ ID NO:322).

This gene is expressed in CD34 positive cells (hematopoietic progenitor cells) and to a lesser extent in adult spleen derived from a chronic lymphocytic leukemia patient.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, chronic lymphocytic leukemia; hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., T-cells and other blood cells, bone marrow, hematopoietic cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Further, nucleic acids and polypeptides of the present invention are useful both diagnostically and therapeutically in the intervention of immune and other disorders in which the ability to alter IL-2 expression is desired. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:201 as residues: Ala-17 to Lys-37, Val-39 to Ser-45, Lys-59 to His-70, Arg-90 to Leu-95, Lys-97 to Lys-107, Ser-117 to Leu-124, Phe-133 to Ser-138, Trp-146 to Leu-167, Pro-175 to Asn-185, Lys-190 to Ser-211, Pro-213 to Ser-222, His-230 to Pro-235, Pro-240 to Pro-246, Pro-253 to Gly-261, Leu-271 to Leu-303, Leu-305 to Leu-326, Lys-343 to Leu-349, Thr-363 to Leu-371, Arg-373 to Tyr-381, Tyr-391 to Leu-401, Pro-404 to Val-414, Ser-426 to Ser-432, Ile-448 to Ser-457, Gln-462 to Trp-468, Lys-477 to Ser-501, Asp-518 to Ser-523, Ala-541 to Gln-554.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of a variety of hematopoietic disorders. The noted expression of this gene in the hematopoietic progenitor cell compartment - as determined by its expression on CD34 positive hematopoietic stem and progenitor cells - indicates that it plays a critical role in the expansion or proliferation of hematopoietic stem/progenitor cells, as well as in the differentiation of the various blood cell lineages. Thus it could be useful in the reconstitution of the hematopoietic system of patients with leukemias and other hematopoietic diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is homologous to heparin cofactor II (HCII) which is a 66-kDa plasma glycoprotein that inhibits thrombin rapidly in the presence of dermatan sulfate or heparin.

This gene is expressed in apoptotic and anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thrombopienia T-cell lymphomas; Hodgkin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system - most notably the T-cell compartment, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The homology to heparin cofactor II (HCII) and the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic disorders particularly in thrombopoesis, most notably of the T-cell compartment. This could include immune modulation, inflammation, immune surveillance, graft rejection, and autoimmunity.

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

The translation product of this gene shares sequence homology with a mouse

protein believed to represent an integral membrane protein.

This gene is expressed in fetal cochlea and epididymus and to a lesser extent in adult spleen and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoclastoma; disorders of the inner ear; male fertility disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inner ear; male reproductive tract; bone; and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cochlea, epididymus and other reproductive tissue, spleen, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:203 as residues: Lys-13 to Gly-23, Cys-38 to Asp-43, Gly-48 to Trp-53, Cys-223 to Ile-237, Ile-240 to Ser-246.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of hearing and fertility disorders. Likewise, it may have a role in the modulation of immune function and in the treatment of osteoporosis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

The translation product of this gene shares sequence homology with reticulocalbin which is thought to be important in the binding of calcium, particularly within the endoplasmic reticulum.

This gene is expressed in endothelial cells and stromal cells and to a lesser extent in osteoblasts, osteoclasts, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoperosis; osteoclastomas; T-cell lymphomas; Hodgkin's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature, bone, and immune systems - particularly the T-cell compartments, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, stromal cells, bone, T-cells and other blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:204 as residues: Lys-20 to Arg-27, Pro-32 to Asp-48, Leu-64 to Arg-72, Asp-108 to Lys-114, Glu-128 to Thr-133, Asp-139 to Phe-147, Thr-196 to Ala-204, Tyr-218 to Glu-228, Val-230 to Gln-236, Arg-241 to Lys-255, Glu-276 to Lys-287.

The tissue distribution and homology to reticulocalbin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of bone disorders such as osteoporosis; the diagnosis and treatment of T-cell lymphomas and Hodgkin's lymphoma; and the treatment of diseases and defects of the vasculature, such as vascular leak syndrome and aberrant angiogenesis that accompanies tumor growth.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

The translation product of this gene shares sequence homology with a family of peptide transport genes - particularly the AtPTR2-B gene from *Arabidopsis* - which are thought to be important in the uptake of small peptides.

This gene is expressed in a number of fetal tissues, most notably lung, brain, cochlea, and liver/spleen, and to a lesser extent in osteoclastoma and endometrial tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoclastoma; endometrial tumors; cancer; leukemias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and endometrium. expression of this gene at significantly higher or lower levels may be

routinely detected in certain tissues (e.g., fetal tissue, pulmonary tissue, bone, brain and other tissue of the nervous system, cochlea, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:205 as residues: Lys-186 to Asn-199, Pro-202 to Ala-207.

The tissue distribution and homology to peptide transport genes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the control of cell proliferation, owing to its strong expression in fetal tissues undergoing active cell division, as well as its expression in a variety of tumors or cancers of adult tissues. Potentially, it may regulate the uptake of peptides that stimulate cell proliferation. This gene product may also be useful in stimulating the uptake of a variety of peptide-based drug compounds.

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

This gene is expressed in fetal liver and spleen and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and tumors of a hematopoietic and/or endothelial cell origin; leukemias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and/or vasculature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, endothelial cells, vascular tissue, and tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:206 as residues: Met-1 to Asp-9, Arg-66 to Gly-76, Asp-164 to Arg-171.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for the treatment of disorders of the immune system. Expression of this gene product in both fetal liver/spleen and endothelial cells indicates that it may be expressed in the hemangioblast, the progenitor cell for both the immune system and the vasculature. Thus, it is most likely expressed in hematopoietic stem cells, and may be useful for the expansion of hematopoietic stem and progenitor cells in conjunction with cancer treatment for a variety of leukemias.

FEATURES OF PROTEIN ENCODED BY GENE NO: 84

The translation product of this gene shares sequence homology with NADH dehydrogenase which is thought to be important in cellular metabolism.

This gene is expressed in fetal dura mater and to a lesser extent in T-cells and hypothalamus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases affecting cellular metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, T-cells and other blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:208 as residues: Pro-27 to Gln-32, Arg-42 to Glu-51.

The tissue distribution and homology to NADH dehydrogenase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases involving cellular metabolism.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene shares sequence homology with I-TRAF, a novel TNF receptor associated factor (TRAF)-interacting protein that regulates TNF receptor-mediated signal transduction. This protein is thought to be important in regulating the cellular response to tumor necrosis factor (TNF), which is an important mediator of inflammation.

This gene is expressed in endothelial cells and to a lesser extent in glioblastoma and osteoblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation; glioblastoma and osteoblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, bone, and glial cells and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:209 as residues: Glu-15 to Thr-22, Glu-46 to Leu-62, Arg-103 to Glu-119, Gln-127 to Glu-132, Asn-152 to Trp-158, Gln-191 to Gln-210, Glu-264 to Thr-271, Tyr-282 to Leu-288, Trp-319 to Thr-331, Glu-335 to Ser-348, Ser-353 to Ser-358, Asp-382 to Asn-392.

The tissue distribution and homology to I-TRAF indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, including rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, particularly where tumor necrosis factor is known to be involved.

FEATURES OF PROTEIN ENCODED BY GENE NO: 86

This gene has homology with a candidate gene involved in X-linked Retinopathy reported by Wong and colleagues (Genomics 15:467-471 (1993)).

This gene is expressed in a T-cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and autoimmune diseases; T-cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory disorders such as sepsis, inflammatory bowel disease, psoriasis, and rheumatoid arthritis as well as autoimmune disease such as lupus. It could also be useful in immune modulation and in the process of immune surveillance. The present invention can be used diagnostically and therapeutically to treat X-linked Retinopathy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

This gene is expressed in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain disorders; neurodegenerative disorders; tumors of a brain origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:211 as residues: Cys-32 to Tyr-38.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of CNS disorders such as epilepsy, paranoia, depression, Alzheimer's disease, and schizophrenia. It could be useful in the survival and/or proliferation of neurons and could effect neuronal regeneration.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HAGEW82	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	11	1679	247	1607	353	353	125	1			30
2	HAGFY16	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	12	1830	87	1786	128	128	126	1	26	27	44
2	HBMCF37	xxxxx 03/19/98	pBluescript	98	1487	79	1487	170	170	212	1	44	45	69
2	HFLQB16	209641 02/25/98	Uni-ZAP XR	99	1653	394	1637	413	413	213	1	25	26	81
3	HALAA60	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	13	1212	1	1212	99	99	127	1	24	25	38
4	HAPBL78	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	14	2061	882	2061	900	900	128	1	22	23	22
5	HASAV70	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	15	1412	10	733	103	103	129	1	20	21	109

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
6	HBNAF22	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	16	1052	276	880	538	538	130	1	17	18	62
7	HBNBL77	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	17	683	1	683	181	181	131	1			29
8	HCDDR90	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	18	1054	86	1007	86	86	132	1	23	24	52
9	HCEEF50	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	19	1393	132	1393	192	192	133	1	17	18	56
10	HCEMU42	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	20	1215	277	1070	401	401	134	1	18	19	215
11	HCENE16	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	21	2042	614	2011	793	793	135	1	26	27	48

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HMSJJ74	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	22	1872	21	1872	69	69	136	1	23	24	67
13	HCUBF15	97923 03/07/97 209071 05/22/97	ZAP Express	23	289	1	289	89	89	137	1	29	30	51
14	HE2DE47	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	24	3533	2821	3532	808	808	138	1	30	31	539
14	HE2DE47	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	100	1145	435	1115	515	515	214	1	22	23	80
15	HKMLH01	209179 07/24/97	pBluescript	25	1148	171	907	196	196	139	1	26	27	56
15	HE6DG34	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	101	734	25	734	295	295	215	1	36	37	48
16	HE9DG49	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	26	717	1	717	70	70	140	1	27	28	200

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
16	HE9DG49	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	102	713	17	713	78	78	216	1	28	29	202
17	HELBA06	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	27	1099	1	1099	38	38	141	1	22	23	215
17	HELBA06	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	103	1080	1	1080	149	149	217	1	25	26	185
18	HSLFM29	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	28	941	171	941	128	128	142	1	42	43	101
19	HELBW38	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	29	756	62	756	294	294	143	1	30	31	111
20	HETHN28	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	30	2100	408	2093	496	496	144	1			19

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
21	HFCDK17	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	31	1448	475	1392	567	567	145	1			29
22	HFEAF41	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	32	456	1	409	21	21	146	1	28	29	98
23	HFKFL13	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	33	1326	1	1322	210	210	147	1			7
24	HFSBG13	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	34	710	1	710	242	242	148	1	16	17	38
25	HFTBE43	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	35	1188	110	1161	178	178	149	1	26	27	130
26	HFTDJ36	97923 03/07/97 209071 05/22/97	Uni-ZAP XR	36	956	1	938	144	144	150	1	21	22	31
27	HKTAC77	97924 03/07/97	Uni-ZAP XR	37	1603	974	1581	1104	1104	151	1			13

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
28	HLHSH36	97924 03/07/97	pBluescript	38	1089	55	1067		209	152	1			7
29	HLHSV96	97924 03/07/97	pBluescript	39	629	1	629	119	119	153	1	32	33	67
30	HLQBQ86	97924 03/07/97	Lambda ZAP II	40	1964	408	1793	581	581	154	1			25
31	HLTBX31	97924 03/07/97	Uni-ZAP XR	41	1522	13	1123	126	126	155	1	32	33	194
32	HLTCJ63	97924 03/07/97	Uni-ZAP XR	42	875	1	875	43	43	156	1	18	19	90
33	HMKAH44	97924 03/07/97	pSport1	43	843	1	843	171	171	157	1	30	31	30
34	HMQAJ64	97924 03/07/97	Uni-ZAP XR	44	489	3	489	55	55	158	1	19	20	89
34	HMQAJ64	97924 03/07/97	Uni-ZAP XR	104	489	6	489	58	58	218	1	22	23	89
35	HOABG65	97924 03/07/97	Uni-ZAP XR	45	534	1	534	17	17	159	1	18	19	88
36	HODCL36	97924 03/07/97	Uni-ZAP XR	46	1374	1	1374	15	15	160	1	20	21	173
36	HODCL36	97924 03/07/97	Uni-ZAP XR	105	640	58	640	72	72	219	1	20	21	137
36	HODCL36	97924 03/07/97	Uni-ZAP XR	106	1529	40	1399	54	54	220	1	27	28	47
37	HODCL50	97924 03/07/97	Uni-ZAP XR	47	596	1	596	269	269	161	1	27	28	44

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
38	HODCV74	97924 03/07/97	Uni-ZAP XR	48	851	99	822	170	170	162	1			22
39	HODCZ16	97924 03/07/97	Uni-ZAP XR	49	2020	569	2020	638	638	163	1	17	18	69
40	HTOEU03	97924 03/07/97	Uni-ZAP XR	50	2432	848	2432	99	99	164	1	19	20	322
40	HTOEU03	97924 03/07/97	Uni-ZAP XR	107	2435	849	2435	928	928	221	1	31	32	69
41	HPBCJ74	97924 03/07/97	pBluescript SK-	51	2340	1627	2340	150	150	165	1	60	61	319
41	HPBCJ74	97924 03/07/97	pBluescript SK-	108	805	92	791	239	239	222	1	21	22	82
42	HPMBU33	97924 03/07/97	Uni-ZAP XR	52	601	188	601	432	432	166	1			30
43	HSAUL66	97924 03/07/97	Uni-ZAP XR	53	359	1	337	142	142	167	1	18	19	71
44	HSIDQ18	97924 03/07/97	Uni-ZAP XR	54	1141	1	1141	25	25	168	1	30	31	280
44	HSIDQ18	97924 03/07/97	Uni-ZAP XR	109	1166	21	1166	433	433	223	1	30	31	42
45	HSJBB37	97924 03/07/97	Uni-ZAP XR	55	1560	63	1148	217	217	169	1			22
46	HSJBQ79	97924 03/07/97	Uni-ZAP XR	56	1507	164	608	57	57	170	1	19	20	326
46	HSJBQ79	97924 03/07/97	Uni-ZAP XR	110	586	4	586	35	35	224	1	23	24	183

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
47	HTEGA76	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	57	450	1	450	83	83	171	1	35	36	68
48	HTEJN13	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	58	1147	1	1147	163	163	172	1	15	16	158
48	HTEJN13	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	111	1134	1	1134	155	155	225	1	19	20	70
49	HTHBL86	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	59	777	1	777	115	115	173	1	18	19	122
50	HTSFO71	97958 03/13/97 209072 05/22/97	pBluescript	60	1191	48	598	52	52	174	1	30	31	128
50	HTSFO71	97958 03/13/97 209072 05/22/97	pBluescript	112	1333	594	1333	829	829	226	1			9
51	HAPNO80	209235 09/04/97	Uni-ZAP XR	61	1580	443	1554	114	114	175	1	1	2	371

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
51	HAUCC47	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	113	1015	249	708	244	244	227	1	28	29	137
52	HBMCL41	97958 03/13/97 209072 05/22/97	pBluescript	62	1117	105	1034	182	182	176	1	28	29	215
53	HCFLD84	97958 03/13/97 209072 05/22/97	pSport1	63	361	1	361	97	97	177	1	32	33	54
54	HE8EM69	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	64	1668	1	1638	150	150	178	1	20	21	22
55	HE8EZ48	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	65	1353	35	1303	231	231	179	1	33	34	102
56	HEBGF73	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	66	1011	655	1011	703	703	180	1	38	39	47

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
57	HFEBF41	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	67	1193	267	1090	459	459	181	1	35	36	95
58	HFRBU14	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	68	560	1	560	63	63	182	1	29	30	94
59	HFVGZ79	97958 03/13/97 209072 05/22/97	pBluescript	69	1657	765	1581	839	839	183	1	21	22	26
60	HHGCM76	97958 03/13/97 209072 05/22/97	Lambda ZAP II	70	711	8	711	270	270	184	1			10
61	HHGCO88	97958 03/13/97 209072 05/22/97	Lambda ZAP II	71	935	111	935	272	272	185	1	19	20	64
62	HHGCP52	97958 03/13/97 209072 05/22/97	Lambda ZAP II	72	504	113	484	127	127	186	1	21	22	21

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
63	HHGDB72	97958 03/13/97 209072 05/22/97	Lambda ZAP II	73	620	1	620	96	96	187	1	18	19	131
64	HHGDI71	97958 03/13/97 209072 05/22/97	Lambda ZAP II	74	581	156	581	248	248	188	1	32	33	68
65	HHSDI45	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	75	1843	537	1786	630	630	189	1	27	28	44
66	HHSEB66	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	76	1441	116	800	167	167	190	1	36	37	64
67	HJPAZ83	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	114	1076	398	1076		575	228	1	11	12	22
68	HLDBO49	97958 03/13/97 209072 05/22/97	pCMVSport 3.0	78	2776	18	1888	187	187	192	1	14	15	169

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
69	HLDBQ19	97958 03/13/97 209072 05/22/97	pCMVSport 3.0	79	1525	401	1480	534	534	193	1	22	23	65
69	HLDBQ19	209226 08/28/97	pCMVSport 3.0	115	1487	401	1487	534	534	229	1	22	23	131
70	HMSGT42	97958 03/13/97 209072 05/22/97	Uni-ZAP XR	80	1563	33	1077	40	40	194	1	32	33	91
71	HMWIC78	97957 03/13/97 209073 05/22/97	Uni-Zap XR	81	1020	18	780	238	238	195	1	23	24	175
72	HTTCT79	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	82	770	101	770	286	286	196	1	26	27	69
73	HNGJU84	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	83	481	1	481	58	58	197	1	20	21	24
74	HNTAC73	97957 03/13/97 209073 05/22/97	pCMVSport 3.0	84	644	1	623	14	14	198	1	25	26	72

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
75	HOSEI45	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	85	1351	435	1284	98	98	199	1	12	13	288
75	HOSEI45	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	116	1350	428	1283		545	230	1			27
76	HOSFD58	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	86	2527	290	1747	56	56	200	1	30	31	623
76	HOSFD58	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	117	2527	288	1747	477	477	231	1	32	33	60
77	HSAUM95	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	87	2566	1843	2566	251	251	201	1	30	31	648
77	HSAUM95	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	118	1098	375	1098	677	677	232	1	21	22	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
78	HSAUR67	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	88	540	1	540	83	83	202	1	32	33	54
79	HSKDI81	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	89	1863	152	1165	188	188	203	1	11	12	265
79	HSKDI81	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	119	1679	152	1166	315	315	233	1			17
80	HSKDW91	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	90	2478	1149	2449	92	92	204	1	19	20	314
81	HTLEX50	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	91	2058	476	2058	414	414	205	1	20	21	206
82	HSKHL65	97957 03/13/97 209073 05/22/97	pBluescript	92	1411	345	1411	157	157	206	1	69	70	194

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
82	HSKHL65	97957 03/13/97 209073 05/22/97	pBluescript	121	1411	345	1411	526	526	235	1	37	38	71
83	HHFGA11	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	93	2187	147	2184	397	397	207	1	30	31	329
83	HHFGA11	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	122	2256	138	2063	228	228	236	1	19	20	95
84	HWTBL40	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	94	757	524	608	445	445	208	1	20	21	57
85	HBXFG80	97957 03/13/97 209073 05/22/97	ZAP Express	95	2394	481	2394	523	523	209	1	1	2	391
86	HCACY32	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	96	672	1	672	117	117	210	1	21	22	25

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
87	HCEDO21	97957 03/13/97 209073 05/22/97	Uni-ZAP XR	97	1419	1	1419	207	207	211	1	20	21	37

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly , the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification , such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown inTable 1, the ORF (open reading frame), or any fragement specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identiy are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the lenght of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignement of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E. coli).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound. such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, or 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any

combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et

al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat

polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined.

First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying

personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell . Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the

proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect

interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes

Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiollitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox , hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteocarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue

regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stoke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotaxic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotaxic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotaxic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

(antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors),or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity , and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, caricadic rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of

positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type

Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

Vector Used to Construct Library	Corresponding Deposited Plasmid
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lafmid BA	plafmid BA
pSport1	pSport1
pCMVSport 2.0	pCMVSport 2.0
pCMVSport 3.0	pCMVSport 3.0
pCR®2.1	pCR®2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR®2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ^{32}P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 µl of reaction mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is

used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P^{32} using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on

either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 ($O.D.^{600}$) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high

affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4° C or frozen at -80° C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 μg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., Virology 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life

Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ^{35}S-methionine and 5 µCi ^{35}S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used

include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No.209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five μg of the expression plasmid pC6 is cotransfected with 0.5 μg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of metothrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μM, 2 μM, 5 μM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 μM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

<u>**Example 9: Protein Fusions**</u>

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion

proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:
GGGATCCGGAGCCCAAATCTTCTGACAAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCCAAAACC
CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC
AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGG
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA
GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC
GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide.

Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl2 (anhyd); 0.00130 mg/L $CuSO_4$-$5H_2O$; 0.050 mg/L of $Fe(NO_3)_3$-$9H_2O$; 0.417 mg/L of $FeSO_4$-$7H_2O$; 311.80 mg/L of Kcl; 28.64 mg/L of $MgCl_2$; 48.84 mg/L of $MgSO_4$; 6995.50 mg/L of NaCl; 2400.0 mg/L of $NaHCO_3$; 62.50 mg/L of NaH_2PO_4-H_20; 71.02 mg/L of Na_2HPO4; .4320 mg/L of $ZnSO_4$-$7H_2O$; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitric Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H_20; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H_20; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H_20; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine: 68.48 mg/ml of L-Phenylalainine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tryrosine-2Na-$2H_20$; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of

Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B_{12}; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in

many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proxial region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

Ligand	tyk2	JAKs Jak1	Jak2	Jak3	STATS	GAS(elements) or ISRE
IFN family						
IFN-a/B	+	+	-	-	1,2,3	ISRE
IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
Il-10	+	?	?	-	1,3	
gp130 family						
IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
Il-11(Pleiotrohic)	?	+	?	?	1,3	
OnM(Pleiotrohic)	?	+	+	?	1,3	
LIF(Pleiotrohic)	?	+	+	?	1,3	
CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
G-CSF(Pleiotrohic)	?	+	?	?	1,3	
IL-12(Pleiotrohic)	+	-	+	+	1,3	
g-C family						
IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
IL-7 (lymphocytes)	-	+	-	+	5	GAS
IL-9 (lymphocytes)	-	+	-	+	5	GAS
IL-13 (lymphocyte)	-	+	?	?	6	GAS
IL-15	?	+	?	+	5	GAS
gp140 family						
IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
IL-5 (myeloid)	-	-	+	-	5	GAS
GM-CSF (myeloid)	-	-	+	-	5	GAS
Growth hormone family						
GH	?	-	+	-	5	
PRL	?	+/-	+	-	1,3,5	
EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
Receptor Tyrosine Kinases						
EGF	?	+	+	-	1,3	GAS (IRF1)
PDGF	?	+	+	-	1,3	
CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is: 5':GCGCCTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCG AAATGATTTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:
5':CTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCGAAATG ATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, Il-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

<u>**Example 13: High-Throughput Screening Assay for T-cell Activity.**</u>

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml genticin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1 x 10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest $2x10e^7$ U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $Na_2HPO_4.7H_2O$, 1 mM $MgCl_2$, and 675 uM $CaCl_2$. Incubate at 37^oC for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37^oC for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting $1x10^8$ cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of $5x10^5$ cells/ml. Plate 200 ul cells per well in the 96-well plate (or $1x10^5$ cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37^oC for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)
5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF-κB (Nuclear Factor κB) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF-κB regulates the expression of genes involved in immune cell activation, control of apoptosis (NF-κB appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κB is retained in the cytoplasm with I-κB (Inhibitor κB). However, upon stimulation, I- κB is phosphorylated and degraded, causing NF- κB to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κB include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF-κB promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF-kB would be useful in treating

diseases. For example, inhibitors of NF-κB could be used to treat those diseases related to the acute or chronic activation of NF-kB, such as rheumatoid arthritis.

To construct a vector containing the NF-κB promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF-κB binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:
5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGACTTTCC
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT
AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC
CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF-κB/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF-κB/SV40/SEAP cassette is removed from the above NF-κB/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF-κB/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF-κB/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1,1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μl of 2.5x dilution buffer into Optiplates containing 35 μl of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μl Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μl Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO_2 incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO_2 incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10^6 cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10^6 cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca^{++} concentration.

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boeheringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg_{2+} (5mM ATP/50mM $MgCl_2$), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM $MgCl_2$, 5 mM $MnCl_2$, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mm EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavadin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phospotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or compliment to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 μg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or mirocapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2- hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water. saline, Ringer's solution. and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

(1) GENERAL INFORMATION:

(i) APPLICANT: Human Genome Sciences, Inc. et al.
(ii) TITLE OF INVENTION: 87 Human Secreted Proteins
(iii) NUMBER OF SEQUENCES: 323
(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.
(B) STREET: 9410 Key West Avenue
(C) CITY: Rockville
(D) STATE: Maryland
(E) COUNTRY: USA
(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
(B) COMPUTER: HP Vectra 486/33
(C) OPERATING SYSTEM: MSDOS version 6.2
(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: March 19, 1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: A. Anders Brookes
(B) REGISTRATION NUMBER: 36,373
(C) REFERENCE/DOCKET NUMBER: PZ004PCT

(vi) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (301) 309-8504
(B) TELEFAX: (301) 309-8439

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 733 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCCCAAATCT TCTGACAAAA CTCACACATG CCCACCGTGC CCAGCACCTG 60

AATTCGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCCAAA ACCCAAGGAC ACCCTCATGA 120

TCTCCCGGAC TCCTGAGGTC ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG 180

TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCGCGGG 240

AGGAGCAGTA CAACAGCACG TACCGTGTGG TCAGCGTCCT CACCGTCCTG CACCAGGACT 300

GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCA ACCCCCATCG 360

AGAAAACCAT CTCCAAAGCC AAAGGGCAGC CCCGAGAACC ACAGGTGTAC ACCCTGCCCC 420

CATCCCGGGA TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT 480

ATCCAAGCGA CATCGCCGTG GAGTGGGAGA GCAATGGGCA GCCGGAGAAC AACTACAAGA 540

CCACGCCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCCT CTACAGCAAG CTCACCGTGG 600

ACAAGAGCAG GTGGCAGCAG GGGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC 660

ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC 720

GACTCTAGAG GAT 733

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser
1 5

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCGCCTCGAG ATTTCCCCGA AATCTAGATT TCCCCGAAAT GATTTCCCCG AAATGATTTC 60

CCCGAAATAT CTGCCATCTC AATTAG 86

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GCGGCAAGCT TTTTGCAAAG CCTAGGC 27

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATTT CCCCGAAATC TAGATTTCCC CGAAATGATT TCCCCGAAAT GATTTCCCCG 60

AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC 120

GCCCCTAACT CCGCCCAGTT CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTTTTTAT 180

TTATGCAGAG GCCGAGGCCG CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240

TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGGACTTTC CC 12

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCGGCCTCGA GGGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60

CCATCTCAAT TAG 73

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTTCCCGGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT 60

CAATTAGTCA GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC 120

CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA 180

GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG 240

CTTTTGCAAA AAGCTT 256

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1679 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GCAGCGCACC CGGGCGATCG CTTCACGGAT GCGGACGACG TAGCCATCCT TACCTACGTG 60

AAGGAAAATG CCCGCTCGCC CAGCTCCGTC ACCGGTAACG CCTTGTGGAA AGCGATGGAG 120

AAGAGCTCGC TCACGCAGCA CTCGTGGCAG TCCCTGAAGG ACCGCTACCT CAAGCACCTG 180

CGGGGCCAGG AGCATAAGTA CCTGCTGGGG GACGCGCCGG TGAGCCCCTC CTCCCAGAAG 240

CTCAAGCGGA AGGCGGAGGA GGACCCGGAG GCCGCGGATA GCGGGGAACC ACAGAATAAG 300

AGAACTCCAG ATTTGCCTGA AGAAGAGTAT GTGAAGGAAG AAATCCAGGA GAATGAAGAA 360

GCAGTCAAAA AGATGCTTGT GGAAGCCACC CGGGAGTTTG AGGAGGTTGT GGTGGATGAG 420

AGCCCTCCTG ATTTTGAAAT ACATATAACT ATGTGTGATG ATGATCCACC CACACCTGAG 480

GAAGACTCAG AAACACAGCC TGATGAGGAG GAAGAAGAAG AAGAAGAAAA AGTTTCTCAA 540

CCAGAGGTGG GAGCTGCCAT TAAGATCATT CGGCAGTTAA TGGAGAAGTT TAACTTGGAT 600

CTATCAACAG TTACACAGGC CTTCCTAAAA AATAGTGGTG AGCTGGAGGC TACTTCCGCC 660

TTCTTAGCGT CTGGTCAGAG AGCTGATGGA TATCCCATTT GGTCCCGACA AGATGACATA 720

GATTTGCAAA AAGATGATGA GGATACCAGA GAGGCATTGG TCAAAAAATT TGGTGCTCAG 780

AATGTAGCTC GGAGGATTGA ATTTCGAAAG AAATAATTGG CAAGATAATG AGAAAAGAAA 840

AAAGTCATGG TAGGTGAGGT GGTTAAAAAA AATTGTGACC AATGAACTTT AGAGAGTTCT 900

TGCATTGGAA CTGGCACTTA TTTTCTGACC ATCGCTGCTG TTGCTCTGTG AGTCCTAGAT 960

TTTTGTAGCC AAGCAGAGTT GTAGAGGGGG ATAAAAAGAA AAGAAATTGG ATGTATTTAC 1020

AGCTGTCCTT GAACAAGTAT CAATGTGTTT ATGAAAGGAA GATCTAAATC AGACAGGAGT 1080

TGGTCTACAT AGTAGTAATC CATTGTTGGA ATGGAACCCT TGCTATAGTA GTGACAAAGT 1140

GAAAGGAAAT TTAGGAGGCA TAGGCCATTT CAGGCAGCAT AAGTAATCTC CTGTCCTTTG 1200

GCAGAAGCTC CTTTAGATTG GGATAGATTC CAAATAAAGA ATCTAGAAAT AGGAGAAGAT 1260

TTAATTATGA GGCCTTGAAC ACGGATTATC CCCAAACCCT TGTCATTTCC CCCAGTGAGC 1320

TCTGATTTCT AGACTGCTTT GAAAATGCTG TATTCATTTT GCTAACTTAG TATTTGGGTA 1380

CCCTGCTCTT TGGCTGTTCT TTTTTTGGAG CCCTTCTCAG TCAAGTCTGC CGGATGTCTT 1440

TCTTTACCTA CCCCTCAGTT TTCCTTAAAA CGCGCACACA ACTCTAGAGA GTGTTAAGAA 1500

TAATGTTACT TGGTTAATGT GTTATTTATT GAGTATTGTT TGTGCTAAGC ATTGTGTTAG 1560

ATTTAAAAAA TTAGTGGATT GACTCCACTT TGTTGTGTTG TTTTCATTGT TGAAAATAAA 1620

TATAACTTTG TATTCGAAAA AAAAAAAAAA AAAATNRCTG CGGNCCGACA AGGGAATTC 1679

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1830 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GCGACCGCGC CCTTCAGCTA GCTCGCTCGC TCGCTCTGCT TCCCTGCTGC CGGCTGCGCA 60

TGGCTTNGGC GTTGGCGGCG CTGGCGGCGG TCGAGCNGCC TGCGSAGCCG GTACCAGCAG 120

TTGCAGAATG AAGAAGAGTC TGGAGAACCT GAACAGGCTG CAGGTGATGC TCCTCCACCT 180

TACAGCAGCA TTTCTGCAGA GAGCGCACAT NATTTTGACT ACAAGGATGA GTCTGGGTTT 240

CCAAAGCCCC CATCTTACAA TGTAGCTACA ACACTGCCCA GTTATGATGA AGCGGAGAGG 300

ACCAAGGCTG AAGCTACTAT CCCTTTGGTT CCTGGGAGAG ATGAGGATTT TGTGGGTCGG 360

GATGATTTTG ATGATGCTGA CCAGCTGAGG ATAGGAAATG ATGGGATTTT CATGTTAACT 420

TTTTTCATGG CATTCCTCTT TAACTGGATT GGGTTTTTCC TGTCTTTTTG CCTGACCACT 480

TCAGCTGCAG GAAGGTATGG GGCCATTTCA GGATTTGGTC TCTCTCTAAT TAAATGGATC 540

CTGATTGTCA GGTTTTCCAC CTATTTCCCT GGATATTTTG ATGGTCAGTA CTGGCTCTGG 600

TGGGTGTTCC TTGTTTTAGG CTTTCTCCTG TTTCTCAGAG GATTTATCAA TTATGCAAAA 660

GTTCGGAAGA TGCCAGAAAC TTTCTCAAAT CTCCCCAGGA CCAGAGTTCT CTTTATTTAT 720

TAAAGATGTT TTCTGGCAAA GGCCTTCCTG CATTTATGAA TTCTCTCTCA AGAAGCAAGA 780

GAACACCTGC AGGAAGTGAA TCAAGATGCA GAACACAGAG GAATAATCAC CTGCTTTAAA 840

AAAATAAAGT ACTGTTGAAA AGATCATTTC TCTCTATTTG TTCCTAGGTG TAAAATTTTA 900

ATAGTTAATG CAGAATTCTG TAATCATTGA ATCATTAGTG GTTAATGTTT GAAAAAGCTC 960

TTGCAATCAA GTCTGTGATG TATTAATAAT GCCTTATATA TTGTTTGTAG TCATTTTAAG 1020

TAGCATGAGC CATGTCCCTG TAGTCGGTAG GGGGCAGTCT TGCTTTATTC ATCCTCCATC 1080

TCAAAATGAA CTTGGAATTA AATATTGTAA GATATGTATA ATGCTGGCCA TTTTAAAGGG 1140

GTTTTCTCAA AAGTTAAACT TTTGTTATGA CTGTGTTTTT GCACATAATC CATATTTGCT 1200

GTTCAAGTTA ATCTAGAAAT TTATTCAATT CTGTATGAAC ACCTGGAAGC AAAATCATAG 1260

TGCAAAAATA CATTTAAGGT GTGGTCAAAA ATAAGTCTTT AATTGGTAAA TAATAAGCAT 1320

TAATTTTTTA TAGCCTGTAT TCACAATTCT GCGGTACCTT ATTGTACCTA AGGGATTCTA 1380

AAGGTGTTGT CACTGTATAA AACAGAAAGC ACTAGGATAC AAATGAAGCT TAATTACTAA 1440

AATGTAATTC TTGACACTCT TTCTATAATT AGCGTTCTTC ACCCCCACCC CCACCCCCAC 1500

CCCCCTTATT TTCCTTTTGT CTCCTGGTGA TTAGGCCAAA GTCTGGGAGT AAGGAGAGGA 1560

TTAGGTACTT AGGAGCAAAG AAAGAAGTAG CTTGGAACTT TTGAGATGAT CCCTAACATA 1620

CTGTACTACT TGCTTTTACA ATGTGTTAGC AGAAACCAGT GGGTTATAAT GTAGAATGAT 1680

GTGCTTTCTG CCCAAGTGGT AATTCATCTT GGTTTGCTAT GTTAAAACTG TAAATACAAC 1740

AGAACATTAA TAAATATCTC TTGTGTAGCA CCTTTTAAAA AAAAAAAAAA AAAAAAAAAA 1800

AAAAAAAAAA AANCCCGGGG GGGGGCCCCN 1830

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1212 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TGTTTGAAGT TGTTACTTTT GTTTACAGCA AAGTTTGATG TAGTGTGCAG TAGTGAGCTC 60

TAGACTGATC TTTTTCTAAA TCAGAAAGTG ATTAAAGTAT GCACAACCAA AGGCAGGTTT 120

TTCTTTTTCA TTTATTCAGC AACTATTTAT TAAGCATCAA CTCTGTGCCA GGCACGTTAC 180

TAGCTGCTAC ATACTGTCTG AACATGACAT ACGGTTAAGT AACTTTACAA TTATTATCAA 240

ATACTTCAAT GTAGATATTT CTTAAGTTGA AATAGCATTA ACTAGGATAA TGCTTTCATG 300

TTATTTTATT TGTCTTGTGA TAGAAATTCA ACTTTGTACC ATCTTAAAAC TAGGTTGCTA 360

TAAAAATAGG AGGATGAAGT CAATAAAGTT TATGCCAGTT TAAAAACTGG AAGGAAAAGG 420

TAAGAGCTCT CCATTATAAA ATAGTTGCAT TCGGTTAATT TTTACACATT AGTGCATTGC 480

GTATATCAAC TGGCCCTCAA TGAAGCATTT AAGTGCTTGG AATTTTACTA AACTGACTTT 540

TTTGCAACTT TGGGAGATTT TTGAGGGGAG TGTTGAAAAT TGCCAAACAC TCACCTCTTA 600

CTCAAAACTT CAAATAAAAT ACACATTTTC AAGAGGGAGC ACCTTTTATA TTTGATAAGT 660

TTTCATTATA AACCTTATAA TACCAGTCAC AAAGAGGTTG TCTGTCTATG GTTTAGCAAA 720

CATTTGCTTT TCTTTTTGGA AGTGTGATTG CAATTGCAGA ACAGAAAGTG AGAAAACACT 780

GCCAGCGGTG ATTGCTACTT GAGGTAGTTT TTTACAACTA CCATTTCCCC TCCATGAAAT 840

TATGTGAAAT TTATTTTATC TTTGGGAAAA GTTGAGAAGA TAGTAAAAGA ATTAGGAATT 900

TAAAATTACA GGGAAAAATA TGTAAGTGAA AAGCAATAAA TATTTTGTTC ACTTTGCTAT 960

CAAGATGTTC ACTATCAGAT ATTTATTATA TGGCAGCAAT TTATATTTTT AATCATTGCC 1020

CATTAATAGA CGCAGTAAAA TATTTTTGAA TCAGACATTT GGGGTTTGTA TGTGCATTAA 1080

AATTGTCTTT TGTACTGTAA GTTACTGTTA ATTTGAATAT TTTATTGAAC TGTCTCCCTG 1140

TGCCTTTATA ATATAAAGTT GTTTCTACAA CTTTTAATGA TCTTAATAAA GAATACTTTA 1200

AGAAAAAAAA AA 1212

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2061 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GGTTTTCCTC CGACTTCCGG ACATCTCCCT GGGAGTCGCG CAGAGTGGAG TCAAAGGCAA 60

CCAGTGCTCG CTGCGGTCTC TGGGGATCGG GACCGCGGCG GCGGCCCGCG AGCGGGATGT 120

TCCGGGGCTT GAGCAGTTGG TTGGGCTTGC AGCAGCCGGT GGCAGGCGGT GGGCAGCCCA 180

ATGGAGATGC TCCACCCGAG CAGCCGTCCG AGACGGTGGC TGAGTCTGCG GAGGAGGAGC 240

TGCAGCAAGC GGGAGACCAG GAGCTCCTCC ACCAGGCCAA AGACTTCGGC AACTATTTAT 300

TTAACTTTGC ATCTGCTGCC ACAAAAAAGA TAACTGAATC AGTTGCTGAA ACAGCACAAA 360

CAATAAAGAA ATCCGTAGAA GAAGGAAAAA TAGATGGCAT CATTGACAAG ACAATTATAG 420

GAGATTTTCA GAAGGAACAG AAAAAATTTG TTGAAGAGCA ACATACAAAG AAGTCAGAAG 480

CAGCTGTGCC CCCATGGGTT GACACTAACG ATGAAGAAAC AATTCAACAA CAAATTTTGG 540

CCTTATCAGC TGACAAGAGG AATTTCCTTC GTGACCCTCC GGCTGGCGTG CAATTTAATT 600

TCGACTTTGA TCAGATGTAC CCCGTGGCCC TGGTCATGCT CCAGGAGGAT GAGCTGCTAR 660

CAAGATGAGA TTTGCCCTCG TTCCTAAACT TGTGAAGGAA GAAGTGTTCT GGAGGAACTA 720

CTTTTACCGC GTCTCCCTGA TTAAGCAGTC AGCCCAGCTC ACGGCCCTGG CTGCCCAACA 780

GCAGGCCGCA GGGAAGGGAG GAGAAGAGCA ATGGCAGAGA GCAAGATTTG CCGCTGGAGA 840

GGCAGTACGG CCCAAAACGC CACCCGTTGT AATCAAATCT CAGCTTAAAA CTCAAGAGGA 900

TGAGGAAGAA ATTTCTACTA GCCCAGGTGT TTCTGAGTTT GTCAGTGATG CCTTCGATGC 960

CTGTAACCTA AATCAGGAAG ATCTAAGGAA AGAAATGGAG CAACTAGTGC TTGACAAAAA 1020

GCAAGAGGAG ACAGCCGTAC TGGAAGAGGA TTCTGCAGAT TGGGAAAAAG AACTGCAGCA 1080

GGAACTTCAA GAATATGAAG TGGTGACAGA ATCTGAAAAA CGAGATGAAA ACTGGGATAA 1140

GGAAATAGAG AAAATGCTTC AAGAGGAAAA TTAGCTGTTC CTGAAATAGA AGAATAATCC 1200

TTAACAGTCT GCAAACTGAC ATTAAATTCT AGATGTTGAC AATTACTGAA TCAGAAGGCA 1260

TGAAAGAGTA TAATTTTATG AAATTCAAAA TTATTCTTTT TTCAAGTTGA AACTTGCCTC 1320

TTCTACTTTA AAAAAGTATA TAGAACAGTT ACTTCTAATA ATCAGAAAGA GATGTTTTAT 1380

AGAACATTTC TTTAATATAA AGTTAGAGAT GTCTTCATAG GCAGTATGGC TATCTTTGCC 1440

ACAGAAACAT AAGTAAAATT TTAGAGTTCT GTTTTCCATG AGGTCAAAAA TATAATTTAT 1500

TCCTCAGTCA TGGTTTTCTA AATATCTGTA CTCCACATTC CATTTTAATT GATATGAGGG 1560

TGTTAAAGTA CCTACTTAAT GGGTTGATTA CTATCAAAAT GACCAAATTA TACCAAAGAA 1620

CTTAAGAGGA AGCACTTTCA GAACTATTCA CTTGCCAGGT ATTTTCTAAA ATTCCACCTG 1680

AAAGCCAAAA GATAAAATAC ATNAGTTGGA TTTTAATGAT ATAAGCATCA CACAATTTTA 1740

CATTAAGAAA TACTGTGCAG CCCATGCGTG GTGGCTCAGG CCTGTAATCC CAGCANTTTG 1800

GGAGGCCGAG GTGGGCAGAT CACCGGAGGT CAGGAGTTCG AGACCAGCCT TGCCAACATA 1860

GTGAAACCCT GTCTTTACTA AAAATACAAA AATTAGCCGG GCATGGTGGC AGGCACCTGT 1920

AATCCCAGCT ACTAGGGAGG CTTTTGAACC CAGGAGGCAG AGGTTGCAGC GAGCTGAGAT 1980

CGCGCCACTG CACTCCAGCC TGGGTGATAG AGTGAGATTC AGTCTCAAAA AAAAAAAAAA 2040

AAAAAAAAAA AATGACCTCG A 2061

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1412 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCCTTCATCT GCGTTGCCAG GAACCCTGTC AGCAGAAACT TCTCAAGCCC CATCCTTGCC 60

AGGAAGCTCT GTGAAGGTGC TGCTGATGAC CCAGATTCCT CCATGGTCCT CCTGTGTCTC 120

CTGTTGGTGC CCCTCCTGCT CAGTCTCTTT GTACTGGGGC TATTTCTTTG GTTTCTGAAG 180

AGAGAGAGAC AAGAAGAGTA CATTGAAGAG AAGAAGAGAG TGGACATTTG TCGGGAAACT 240

CCTAACATAT GCCCCCATTC TGGAGAGAAC ACAGAGTACG ACACAATCCC TCACACTAAT 300

AGAACAATCC TAAAGGAAGA TCCAGCAAAT ACGGTTTACT CCACTGTGGA AATACCGAAA 360

AAGATGGAAA ATCCCCACTC ACTGCTCACG ATGCCAGACA CACCAAGGCT ATTTGCCTAT 420

GAGAATGTTA TCTAGACAGC AGTGCACTCC CCTAAGTCTC TGCTCAAAAA AAAAACAATT 480

CTCGGCCCAA AGAAAACAAT CAGAAGAATT CACTGATTTG ACTAGAAACA TCAAGGAAGA 540

ATGAAGAACG TTGACTTTTT TCCAGGATAA ATTATCTCTG ATGCTTCTTT AGATTTAAGA 600

GTTCATAATT CCATCCACTG CTGAGAAATC TCCTCAAACC CAGAAGGTTT AATCACTTCA 660

TCCCAAAAAT GGGATTGTGA ATGTCAGCAA ACCATAAAAA AAGTGCTTAG AAGTATTCCT 720

ATAAAAATGT AAATGCAAGG TCACACATAT TAATGACAGC CTGTTGTATT AATGATGGCT 780

CCAGGTCAGT GTCTGGAGTT TCATTCCATC CCAGGGCTTG GATGTCAGGA TTATACCAAG 840

AGTCTTGCTA CCAGGAGGGC AAGAAGACCA AAACAGACAG ACAAGTCCAG CAGAAGCAGA 900

TGCACCTGAC AAAAATGGAT GTATTAATTG GCTCTATAAA CTATGTGCCC AGCAYTATGC 960

TGAGCTTACA CTAATTGGTC AGACATGCTG TCTGCCCTCA TGAAATTGGC TCCAAATGAW 1020

TGAACTACTT TCATGAGCAG TTGTAGCAGG CCTGACCACA GATTCCCAGA GGGCCAGGTG 1080

TGGATCCACA GGACTTGAAG GTCAAAGTTC ACAAAGATGA AGAATCAGGG TAGCTGACCA 1140

TGTTTGGCAG ATACTATAAT GGAGACACAG AAGTGTGCAT GGCCCAAGGA CAAGGACCTC 1200

CAGCCAGGCT TCATTTATGC ACTTGTCTGC AAAAGAAAAG TCTAGGTTTT AAGGCTGTGC 1260

CAGAACCCAT CCCAATAAAG AGACCGAGTC TGAAGTCACA TTGTAAATCT AGTGTAGGAG 1320

ACTTGGAGTC AGGCAGTGAG ACTGGTGGGG CACGGGGGGC ANTGGGTANT GTAAACCTTT 1380

TAAAGATGGT TAATTCNTCA TTAGTGTTTT TT 1412

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1052 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TTCCTCTCCT CTCTCTACCC CTCCTGTCTC TCCTCCCCTC CTCTCTCTTC CTCTCCTCTC 60

TCTCTTCCTC TCCTCTCTCT TCCCTTCCTG TCTCTCTTCC CCTCCTCTCT CTCTTCCTGT 120

CCTCTATCTC TTCCCCTCCT CTATCTCTTC CTCTCCTCTC TCTCTTCCTC TCCTCTCTCT 180

CTCTTSCTTT CTTCTCTCTC TCCTGTCTCG GCTGTTGTGG GTTGCAGGTT GGGTGCTGCT 240

GTTGTGGTCC TTCCCAGAAA CTGCCAGTAG AGGGCAGCCT GGGCATCCTA ATGCTTACTC 300

TGGTTGTTAC ACAAAGAAAA TATTGGGGTC ACTGGCGAGC CCACCCACAC TCACCAGAAT 360

CTCCACTGTA GTCCCCCTAA CAAACAGCCC TTCACTTCCT CTCCCACTTC AGCAATTTGT 420

ATTTTGATGC CATTGGCCTC AGATCAGAGT GTTTTAAATC ATCACGCCCT GGCTTATCCC 480

TGGTCGAGCC AGGACACGGG GTGCTTCAGT GGGTCTGTCA CCCTCTCTCC TTGAAGCATG 540

TTGCTTTTAT TTATTTACTT TTACTCTCAC CCTGCTCCTG TACCAGCAGG GGCCACTTCA 600

AAGCCAAGGT ACAGGGTGAT AACTTGTGGT CCAGCATCAG TTTTCTCCAC TTCTTTCTCC 660

CACTCACCCC CAGCAAGGTG CCTGGGGAGA CTTGAGCAGA TGTTTCATTT TGGCCTGGCC 720

AGTGGCTGAA AGCAGGCCTC CAATGCACTG TGACCTCTGG CTTCCCCAGC AGCTTTCCCA 780

GAGAGGCAGA GGGGCCTTCC ACAGCCCGGG TTCTCCTGCT GCCTCCTGCC TGCTGCAGCT 840

GCAGGCATTC TGAGGGGCAA CGTGGAGGAA GGGCCAGGGA TGCATGGGAT TTTAATTGTT 900

TCATCACACC TTCCCCGTGG CAAAGAAACA GTCAGTCCTC TTCAGGTGTC TTCTGGATTT 960

CTGGTGATGG ACAGAGAAAT CTTTTTACAG TTTCAAATTA TGTTCAACAA ATAAAAATTG 1020

CATTTTTTAT TTTGGAAAAA AAAAAAAAAA AA 1052

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 683 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AATTCGGCAG AGGCACTTAT CATGTACATA TAGCCTGTTT TTTAGCATTG TTAGACAAAG 60

TAGGCATATT CCTTTCCATC CAAGAACTCA TAACCTAGTA ATTGTAGTTG GCTGATAGCT 120

CATTGCCCAT ACACAAGGAT CTAACACAAC CTCTTGAATA AACATCCCCC TTATTCAGAA 180

ATGCCTTTTC CTATTTCCAT ATTGCAACTT TGCTTACAAA TTTCCAATCT GTCTTTCTGT 240

TTACAGAAGA TATACAAAAT TCCTTTTGTA TGATCTCTTT ATATCTCTTG ATTTTCTTTT 300

GTGTTTGCTA CCAAAGGGCC TGCACATAGT GAGAAGATTG TGCATGATCT GTGAGCTCTA 360

CCACACCTGG AATTAGGGAT CACCAATATG AGAAAAAAAA TTGGAGGTAC AAATAACATT 420

ATCATATGTW ATTGGCATAT AAATTACAGA TGTWTCTATG ACTAAAAACC CTGTGGATAT 480

WAACCMAATG CAGATAAWTW TAATAAAATW TWTAAAAATW TWATCMAATA ATGATAGTGC 540

TATTCAAATA CTTCAAATTT GCACAGTGAT TTATTTCTTA AAATATGTTA ACACATGTGA 600

GCCAATACAC TGAGGTCACT GGATAAATAA ACAGATTCTT GCAAAAAAAA AAAAAAAAAA 660

ACTCGAGGGG GGCCCGTACC CTT 683

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1054 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AAACTCATTT AGGTGACACT ATAGAAGGTA CGCCTGCAGG TACCGGTCCG GAATTCCCGG 60

GTCGACCCAC GMGNCCGGCG ACAAGATGGC AGCAGCGTGT CGGAGCGTGA AGGGCCTGGT 120

GGCGGTAATA ACCGGAGGAG CCTCGGGCCT GGGCCTGGCC ACGGCGGACG ACTTGTGGGG 180

CAGGGAGCCT CTGCTGTGCT TCTGGACCTG CCCAACTCGG GTGGGGAGGC CCAAGCCAAG 240

AAGTTAGGAA ACAACTGCGT TTTCGCCCCA GCCGACGTGA CCTCTGAGAA GGATGTGCAA 300

ACAGCTCTGG CTCTAGCAAA AGGAAAGTTT GGCCGTGTGG ATGTAGCTGT CAACTGTGCA 360

GGCATCGCGG TGGCTAGCAA GACGTACAAC TTAAAGAAGG GCCAGACCCA TACCTTGGAA 420

GACTTCCAGC GAGTTCTTGA TGTGAATCTC ATGGGCACCT TCAATGTGAT CCGCCTGGTG 480

GCTGGTGAGA TGGGCCAGAA TGAACCAGAC CAGGGAGGCC AACGTGGGGT CATCATCAAC 540

ACTGCCAGTG TGGCTGCCTT CGAGGGTCAG GTTGGACAAG CTGCATACTC TGCTTCCAAG 600

GGGGGAATAG TGGGCATGAC ACTGCCCATT GCTCGGGATC TGGCTCCCAT AGGTATCCGG 660

GTGATGACCA TTGCCCCAGG TCTGTTTGGC ACCCCACTGC TGACCAGCCT CCCAGAGAAA 720

GTGTGCAACT TCTTGGCCAG CCAAGTGCCC TTCCCTAGCC GACTGGGTGA CCCTGCTGAG 780

TATGCTCACC TCGTACAGGC CATCATCGAG AACCCATTCC TCAATGGAGA GGTCATCCGG 840

CTGGATGGGG CCATTCGTAT GCAGCCTTGA AGGGAGAAGG CAGAGAAAAC ACACGCTCCT 900

CTGCCCTTCC TTTCCCTGGG GTACTACTCT CCAGCTTGGG AGGAAGCCCA GTAGCCATTT 960

TGTAACTGCC TACCAGTCGC CCTCTGTGCC TAATAAAGTC TCTTTTTCTC ACANAAAAAA 1020

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA 1054

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1393 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

```
GGAACAAGCT GGGATATGTG AGCGTTAAGC TACTCACATC CTTCAAAAAG GTGAAACATC     60
TTACACGGGA CTGGAGAACC ACAGCACATG CTTTGAAGTA TTCAGTGGTC CTTGAGTTGA    120
ATGAGGNCCA CCGGAAGGTG AGGAGGACCA CCCCCGTCCC ACTGTTCCCC AACGAGAACC    180
TCCCCAGCAA GATGCTCCTG GTCTATGATC TCTACTTGTY TCCTAAGCTG TGGGCTCTGG    240
CCACCCCCCA GAAGAATGGG AAGGGTGCAA GARAAGGTGA TGGAACACCT GCTCAAGCTT    300
TTTGGGACTT TTGGAGTCAT CTCATCAGTG CGGATCCTCA AACCTGGGAG AGAGCTGCCC    360
CCTGACATCC GGAGGNTCCA GCAGCCGCTA CAGCTCCTCT GACCCCGAGA GCAACCCCAC    420
ATCCCCTATG GCGGGCCGAC GGCACGNGKC CACCAACAAG CTCAGCCCGT CTGGCCACCA    480
GAATCTCTTT CTGAGTCCAA ATGCCTCCCC GTGCACAAGT CCTTGGAGCA GCCCCTTGGC    540
CCAACGCAAA GGCGTTTCCA GAAAGTCCCC ACTGGCGGAG GAAGGTAGAC TGAACTGCAG    600
CACCAGCCCT GAGATCTTCC GCAAGTGTAT GGATTATTCC TCTGACAGCA GCGTCACTCC    660
CTCTGGCAGC CCCTGGGTCC GGAGGCGTCG CCAAGCCGAG ATGGGGACCC AGGAGAAAAG    720
CCCCGGTACG AGTCCCCTGC TCTCCCGGAA GATGCAGACT GCAGATGGGS TACCCGTAGG    780
TNGCTTGAGG TTGCCCAGGG GTCCTGACAA CACCAGAGGA TTTCATGGCC ATGAGAGGAG    840
CAGGGCCTGT GTATAAATAC CTTCTATTTT TAATACAAGC TCCACTGAAA ACCACCTTCG    900
TTTTCAAGGT TCTGACAAAC ACCTGGCATG ACAGAATGGA ATTCGTTCCC CTTTGAGAGA    960
TTTTTTATTC ATGTAGACCT CTTAATTTAT CTATCTGTAA TATACATAAA TCGGTACGCC   1020
ATGGTTTGAA GACCACCTTC TAGTTCAGGA CTCCTGTTCT TCCCAGCATG GCCACTATTT   1080
TGATGATGGC TGATGTGTGT GAGTGTGATG GCCCTGAAGG GCTGTAGGAC GGAGGTTCCC   1140
TGGGGGAAGT CTGTTCTTTG GTATGGAATT TTTCTCTCTT CTTTGGTATG GAATTTTTCC   1200
CTTCAGTGAC TGAGCTGTCC TCGATAGGCC ATGCAAGGGC TTCCTGAGAG TTCAGGAAAG   1260
TTCTCTTGTG CAACAGCAAG TAGCTAAGCC TATAGCATGG TGTCTTGTAG GACCAAATCG   1320
ATGTTACCTG TCAAGTAAAT AAATAATAAA ACACCCAACT GGGAGTGCTG AAAAAAAANA   1380
ANNAAAAAAC TCG                                                      1393
```

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1215 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AGGAAAAGTT TTCCNAATTG GAAAGCGGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG 60

NTCANTCATT AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGNTCGTAT GTTGTGTGGA 120

ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCTN 180

TAATACGACT CACTATAGGG AAAGCTGGTA CGCCTGCAGG TACCGGTCCG GAATTCCCGG 240

GTCGACCCAC GCGTCCGCCC ACGCGTCCGT GAAAATCCGA AGTGCCGCGG AAAGTGGAGG 300

TGAGGGCCGC CCGCCCTAGA GGTGCCCGTC CGAGAGGCAG AGCTGACAAG GAAGGTTTCG 360

AGCGTTTTGC TGGCAAAGGG ATTTCTTACA ACCTCCAGGC ATGCGTCTTT CTGCCCTGCT 420

GGCCTTGGCA TCCAAGGTCA CTCTGCCCCC CCATTACCGC TATGGGATGA GCCCCCCAGG 480

CTCTGTTGCA GACAAGAGGA AGAACCCCCC ATGGATCAGG CGGCGCCCAG TGGTTGTGGA 540

ACCCATCTCT GATGAAGACT GGTATCTGTT CTGTGGGGAC ACGGTGGAGA TCCTAGAAGG 600

CAAGGATGCC GGGAAGCAGG GCAAAGTGGT TCAAGTTATC CGGCAGCGAA ACTGGGTGGT 660

CGTGGGAGGG CTGAACACAC ATTACCGCTA CATTGGCAAG ACCATGGATT ACCGGGGAAC 720

CATGATCCCT AGTGAAGCCC CCTTGCTCCA CCGCCAGGTC AAACTTGTGG ATCCTATGGA 780

CAGGAAACCC ACTGAGATCG AGTGGAGATT TACTGAAGCA GGAGAGCGGG TACGAGTCTC 840

CACACGATCA GGGAGAATTA TCCCTAAACC CGAATTTCCC AGAGCTGATG GCATCGTCCC 900

TGAAACGTGG ATTGATGGCC CCAAAGACAC ATCAGTGGAA GATGCTTTAG AAAGAACCTA 960

TGTGCCCTGT CTAAAGACAC TGCAGGAGGA GGTGATGGAG GCCATGGGGA TCAAGGAGAC 1020

CCGGAAATAC AAGAAGGTCT ATTGGTATTG AGCCTGGGGC AGAGCAGCTC CTCCCCAACT 1080

TCTGTCCCAG CCTTGAAGGC TGAGGCACTT CTTTTTCAGA TGCCAATAAA GAGCACTTTA 1140

TGAGTCCTCC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1200

AAAAGGGGCG GCCGC 1215

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2042 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CTGCATCCAG GCGCAGAATA ACCTGGGTAT CTTGTGGTCT GAAAGAGAGA AATTGAAACT 60

GCACAGGCTT ACCTAGAGTC ATCAGAAGCA CTATATAATC AGTATATGAA AGAGGTTGGG 120

AGTCCTCCTC TTGATCCTAC TGAGCGTTTT CTTCTGAAGA AGAGAAACTT ACTGAACAAG 180

AGAGATCAAA AAGATTTGAA AAGGTTTATA CTCATAACCT ATATTACCTA GCTCAAGTCT 240

ACCAGCATCT GGAAATGTTT GAGAAGGCTG CTCACTATTG CCATAGTACA CTAAAACGCC 300

AGCTTGAGCA CAATGCCTAC CATCCTATAG AGTGGGCTAT CAATGCTGCT ACCTTGTCAC 360

AGTTTTACAT CAATAAGCTA TGCTTTATGG AGGCCAGGCA CTGTTTATCA GCTGCTAATG 420

TCATTTTTGG TCAAACTGGA AAGATCTCAG CCACAGAAGA CACTCCTGAA GCTGAAGGAG 480

AAGTGCCAGA GCTTTATCAT CAAAGAAAGG GGGAAATAGC AAGGTGCTGG ATCAAATACT 540

GTTTGACTCT CATGCAGAAT GCCCAACTCT CCATGCAGGA CAACATAGGA GAGCTTGATC 600

TTGATAAACA GTCTGAACTT AGAGCTTTAA GGAAAAAAGA ACTAGATGAG GAGGAAAGCA 660

TTCGGAAAAA AGCTGTGCAG TTTGGAACCG GTGAACTGTG TGATGCCATC TCTGCAGTAG 720

AAGAGAAAGT GAGCTACTTG AGACCTTTAG ATTTTGAAGA AGCCAGAGAA CTTTTCTTAT 780

TGGGTCAGCA CTATGTCTTT GAGGCAAAAG AGTTCTTTCA GATTGATGGT TATGTCACTG 840

ACCATATTGA AGTTGTCCAA GACCACAGTG CTCTGTTTAA GGTGCTTGCA TTCTTTGAAA 900

CTGACATGGA GAGACGGTGC AAGATGCATA AACGCRGAAT AGCCATGCTA GAGCCCCTAA 960

CTGTAGACCT GAATCCACAG TATTATCTGT TGGTCAACAG ACAGATCCAG TTTGAAATTG 1020

CACATGCTTA CTATGATATG ATGGATTTGA AGGTTGCCAT TGCTGACAGG CTAAGGGATC 1080

CTGATTCACA CATTGTAAAA AAAATAAATA ATCTTAATAA GTCAGCACTG AAGTACTACC 1140

AGCTCTTCTT AGACTCCCTG AGAGACCCAA ATAAAGTATT CCCTGAGCAT ATAGGGGAAG 1200

ATGTTCTTCG CCCTGCCATG TTAGCTAAGT TTCGAGTTGC CCGTCTCTAT GGCAAAATCA 1260

TTACTGCAGA TCCCAAGAAA GAGCTGGAAA ATTTGGCAAC ATCATTGGGA ACATTACAAA 1320

TTTATTGTTG ATTACTGTGA AAAGCATCCT GAGGCCGCCC AGGAAATAGA AGTTGAGCTA 1380

GAACTTAGTA AAGAGATGGT TAGTCTTCTC CCAACAAAAA TGGAGAGATT CAGAACCAAG 1440

ATGGCCCTGA CTTAATCCTT GTTTTTAAAG AAAGGAAATG TGCAATATTG AAGTGATCTT 1500

TTTCCCTAGT CAGACAGGCC CAATTCCATT GTGATGTTTA CCTTTATAGC CAGGTGAGTG 1560

CAGTTTGAAC TTGAGATACA GTCAACTGAG TGTTTGCTAG GATCCTAAGG AACATAAAGT 1620

TAATTAAAAA CTTACACCTA ATTATGTAAA TTGCCTTGTT AAAGACATGT GATTTGTATT 1680

TTAGATGCTT GTTTCCTATT AAAATACAGA CATTTCTACC CTCAGTTTCT AAATGTAGAC 1740

TATTTGTTGG CTAGTACTTG ATAGATTCCT TGTAAGAAAA AATGCTGGGT AATGTACCTG 1800

GTAACAAGCC TGTTAATATA TTAAGATTGA AAAAGTAACT TCTATAGTTA CTCCTTCTAA 1860

AATATTTGAC TTCCTACATT CCCCCCACCC AAAATCTTTC CCTTTTGAAA ATACTAAAAA 1920

CTAAGTTATG TTATTATAAA GTGTAAAATG GTTTGTCTTA ATTATAGGAG AAAAAGGCCT 1980

TGTTAGAAAT AAAATAAACT GACTTATTTC ACTAATGAAA AAAAAAAAAA AAAAAAAAAA 2040

TT 2042

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGGTCGACCC ACGCGTCCGA TTGGCCTAGA GCTCCTGTGA CCGAGAGCGC CACGGAAGCC 60

TGGGGATGAT GTCGGGCAGC TTTATTCTTT GCTTGGCTTT GGTAACTAGG TGGTCCCCTC 120

AAGCATCCTC AGTTCCTCTT GCTGTTTATG AATCTAAGAC AAGGAAGTCC TATAGAAGCC 180

AAAGGGACAG GGACGGAAAG GACAGGTCCC AAGGGATGGG GCTGTCTTTA CTTGTGGAAA 240

CCAGGAAATT GCTCCTCTCA GCCAACCAAG GTTGACCACA CACCACCCTT CCGGAGCAGC 300

TCAGTCAGCC CTCGGGGACG RGAAACCACA AGCGCAGAGA CGCTGAGGCC CAGGCAGGTG 360

AAGAGGAAGT GGCTTTGGGT TTTTAAAGTA GGTGAGCGTG ACCTCTCTGA CTGCTTCTTC 420

CCCGGGGGGG ACTGCAAACC GCTCAGGGTT GCGGCAGAGC CATGGACTTC CGGTCCCTGC 480

AACGGGTGAC CTAAGCGTGG TGCACCCATC AGTCACGCAG GAGGACTGAC TTGACAGACG 540

AAAGACAAGC CCGGATGACA CAGGGTGAGA AGAGTCAGGG CCGCACCTCT GTCCCTGCAA 600

ACCAACAGGT GCATGGTGAG TGTGGCAGTC CCCACAGCTC CACAATGGGC TCCCCCGCCA 660

ACGGGGACGA CAGGGATCTT CAGGAACTTC TGACCTCACC AAGTCAAGTG GACCACTCTC 720

CACTCCACGA GGATGTGAAA CGGTTCTTTA AAATGGGATT TTAGAGCCTC GGGAATGCAT 780

GTGCGTCGCA TCTTTCATAT TATGGGTCAG GATAGATTCA TTTCTTGCAA CATAGTGGAA 840

AAGATATAAG CTGCAGTAAT TTGCTCTTTG AATGACCGTC ACCCCCAGTA TAGGATATGC 900

TTGTATCCCC CCGTCACTCC TCCGCCTGTT TTTTAAACTT TTCCACCACC TGCGTCCAAA 960

AAGAATGTTA TAGCGAGTGC TCTTAAATGT TGAACCTGGG TGTTGCTTCC GGGCCAGTCT 1020

GCGTGGCTCC ATGAAAAGCT CACTGCTGCC CCAGCCGGGC TTCTTAGAGG AGGTCAGTTG 1080

TCCTATGTAT CATCATTTAC TCTGGGAATC CTACTGTGAA ATCATGTCTG TATTTTTCTG 1140

GAGCAGTTCA CATAGAGTAG AATGTGGAAT TTCCCGTGAA CGTCTCCTTC CTCCCCCGTA 1200

TCTGCCGCCT GTCACTTCGC CACCGTGCTA GAATACTGTT GTGTTGTAAG ATGACTAATT 1260

TTAAAAGAAC CTGCCCTGAA AAGTTCTTAG AAACGCAATG AAAGGGAGGA ACTTGTCCTT 1320

TACCCAGTTT TTCCTTTGTA GGATGGGAAA GTATAAAAAG GCACAGAAGG TTGTCATGGG 1380

CTGTTCCTTG GGGGTTTTTA TCCTGCTCAC CGTGGAGATA AGCCTGCGGC TTGTCTAACC 1440

AGCGCAGCGM AAAGGTCTCA ATGCCTTTTG GTAACATCCG TCATTGCAGA AGAAAGTTTA 1500

CACGACGTCA AAAAGTGACG TTCATGCTAA GTGTTTTTCC AGAAATATTG GTTTCATGTT 1560

TCTTATTKGC TCTGCCTCCT GTGCTTATAT CATCCAAAAA CTTTTTAAAA AGGTCCAGAA 1620

TTCTATTTTA ACCTGATGTT GAGCACCTTT AAAACGTTCG TATGTGTGTT GCACTAATTC 1680

TAAACTTTGG AGGCATTTTG CTGTGTGAGG CCGATCGCCA CTGTAAAGGT CCTAGAGTTG 1740

CCTGTTTGTC TCTGGAGATG GAATTAAACC AAATAAAGAG CTTCCACTGG AGGCTTGTAT 1800

TGACCTTGTA ACTATATGTT AATCTCGTGT TAAAATAAAA TATAACTTGT GAAAAAAAAA 1860

AAAAAAAAAC NT 1872

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 289 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CATTTACCCA CCTATCAACA TGTTTGCTTT CTCTTTTGTT GGTGAGAATG AGTGGCTTCT 60

TGCTCCTAGC TAGAGCCAGT CCTTCCATAT GTGCTTTAGA TTCTTCCTGT TTTGTTCAAG 120

AATATTGCTC AAGCTATTCT TCCTCCTGTT TCCTGCATCA GCATTTCCCC TCTCTACTAG 180

ATCATCTCTG TCAGTAAATG AACATGTTGT TGTTTCTCCT AGAAGTACTG TTTCTATATC 240

TAGATAGTAC TCTAGCTAGA GTTAAAAAAA AAAAAAAAAA CCTNGGGGG 289

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3533 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

```
TTTTATTTAC TTCAAATTAA CTGTACTTTA CTCAAATAGA AAANGAATAA TTTTCACATT     60
ATGAAGCTAC ACAATTCCAA AATACACATG CTGAGGCTCT TTTTAAGTCC GAATTGTCTA    120
GTAATTACAA AAAAGTGAAG AGTTTACAGA TATACAAGGA AATAAAGGCG AATTATTGCA    180
AAGAAAACAA GTTTAATTTC ACTTTGAATG ACAACGATTT TTCTGGAAAG CAGATACTTC    240
ACTCCTTTAA GTTTCCACCC AAGCCACAAT AATTTCAAAC GGTCTTGCGG ATGACCCAGC    300
TGGTCACTCT TGTTTATGTG GGGACTGGAG GTAATGAGAG CCAAAAAAAG TGCTATAAAC    360
CTAATTTGGC TAGAGCAAGT TCACACGACA CGACCGTGCT TTAAAAACTT GCTCTCCATT    420
ATGTACTTCC TTCCATCAGG TTGGGGAAAA AAAAATGGTG GGGATGGTGA GTAAACACAC    480
CAGTGGTTTC ATCAGAGGGG AACTCACTAC TCAGGAGGTG ACGGTGACGT GGTGCCGGTC    540
CCTGAAGTAC GCGCACAAGC TCCGGAGGTT GCGGGAGCTT CCGCTGCCGC CTGGAGGGAA    600
GCCGGAGCGA CGGGGGTCAC GGCGGCGGTC AGAGGGTAAA GGTCTTGCTC CCAGCAGCCT    660
CCGCGGTGGA TACGTCGCCA TCTTGGATCC GCGGGACAAG AAAATTCATG CGAGGGAGAC    720
GTGGTGGGCG GTCCTTCCTG TGACACGACC CTTGAGTGAC AGTTCTATTT GATTGCCTCC    780
GGTACTGTGA GGAAAGGACA CGACTCTATG GTGAGGACTG ATGGACATAC ATTATCTGAG    840
AAAAGAAACT ACCAGGTGAC AAACAGCATG TTTGGTGCTT CAAGAAAGAA GTTTGTAGAG    900
GGGGTCGACA GTGACTACCA TGACGAAAAC ATGTACTACA GCCAGTCTTC TATGTTTCCA    960
CATCGGTCAG AAAAAGATAT GCTGGCATCA CCATCTACAT CAGGTCAGCT GTCTCAGTTT   1020
GGGGCAAGTT TATACGGGCA ACAAAGTGCA CTAGGCCTTC CAATGAGGGG GATGAGCAAC   1080
AATACCCCTC AGTTAAATCG CAGCTTATCA CAAGGCACTC AGTTACCGAG CCACGTCACG   1140
CCAACAACAG GGGTACCAAC AATGTCACTT CACACGCCTC CATCTCCAAG CAGGGGTATT   1200
TTGCCTATGA ATCCTARGAA TATGATGAAC CACTCCCAGG TTGGTCAGGG CATTGGAATT   1260
CCTAGCAGGA CAAATAGCAT GAGCAGTTCA GGGTTAGGTA GCCCCAACAG AAGCTCGCCA   1320
AGCATAATAT GTATGCCAAA GCAGCAGCCT TCTCGACAGC CTTTTACTGT GAACAGTATG   1380
TCTGGATTTG GAATGAACAG GAATCAGGCA TTTGGAATGA ATAACTCCTT ATCAAGTAAC   1440
```

ATTTTTAATG GAACAGACGG AAGTGAAAAT GTGACAGGAT TGGACCTTTC AGATTTCCCA 1500

GCATTAGCAG ACCGAAACAG GAGGGAAGGA AGTGGTAACC CAACTCCATT AATAAACCCC 1560

TTGGCTGGAA GAGCTCCTTA TGTTGGAATG GTAACAAAAC CAGCAAATGA ACAATCCCAG 1620

GACTTCTCAA TACACAATGA AGATTTTCCA GCATTACCAG GCTCCAGCTA TAAAGATCCA 1680

ACATCAAGTA ATGATGACAG TAAATCTAAT TTGAATACAT CTGGCAAGAC AACTTCAAGT 1740

ACAGATGGAC CCAAATTCCC TGGAGATAAA AGTTCAACAA CACAAAATAA TAACCAGCAG 1800

AAAAAAGGGA TCCAGGTGTT ACCTGATGGT CGGGTTACTA ACATTCCTCA AGGGATGGTG 1860

ACGGACCAAT TTGGAATGAT TGGCCTGTTA ACATTTATCA GGGCAGCAGA GACAGACCCA 1920

GGAATGGTAC ATCTTGCATT AGGAAGTGAC TTAACAACAT TAGGCCTCAA TCTGAACTCT 1980

CCTGAAAATC TCTACCCCAA ATTTGCGTCA CCCTGGGCAT CTTCACCTTG TCGACCTCAA 2040

GACATAGACT TCCATGTTCC ATCTGAGTAC TTAACGAACA TTCACATTAG GGATAAGCTG 2100

GCTGCAATAA AACTTGGCCG ATATGGTGAA GACCTTCTCT TCTATCTCTA TTACATGAAT 2160

GGAGGAGACG TATTACAACT TTTAGCTGCA GTGGAGCTTT TTAACCGTGA TTGGAGATAC 2220

CACAAAGAAG AACGAGTATG GATTACCAGG GCACCAGGCA TGGAGCCAAC AATGAAAACC 2280

AATACCTATG AGAGGGGAAC ATATTACTTC TTTGACTGTC TTAACTGGAG GAAAGTAGCT 2340

AAGGAGTTCC ATCTGGAATA TGACAAATTA GAAGAACGGC CTCACCTGCC ATCCACCTTC 2400

AACTACAACC CTGCTCAGCA AGCCTTCTAA AAAAAAAAAA AAAAAAAAAA AAAAAGACTT 2460

CCCTTTTCTT GGGGTATGGC TGTCTCAGCA CAATACTCAA CATAACTGCA GAACTGATGT 2520

GGCTCAGGCA CCCTGGTTTT AATTCCTTGA GGATCTGGCA ATTGGCTTAC GCAAAAGGTC 2580

ACCATTTGAG GTCCTGCCTT ACTAATTATG TGCTGCCCAA CAACTAAATT TGTAATTTGT 2640

TTTTCTCTAG TTTGAGCAGG GTCTGAATTT TTTCATTTAT TTCCTTTTTT GCCAGCAGAC 2700

AGACTTGAGT CTGTAAAGAC AAGCAAATAC ACTGACAGAA GTTTACCATA GTTTCTAAAA 2760

TGTAAAAAAG AAAACCCCCA AAAGACTCAA GAAAATTAGA CCACAAATTT TGCATTGTTC 2820

ATTGTAGCAC TATTGGTAAT AAAATAACAA ATGTTTGTGC ATTTTTATGT GAAGATCCTT 2880

CTCGTATTTC ATTTGGAAAG ATGAGCAAGA GGTCTGCTTC CTTCATTTTA CTTCCCCTTC 2940

TGTTTTTGAA AGGCAGTTTC GCCAAGCTTA ATGCAAGAAT ATCTGACTGT TTAGAAGAAA 3000

GATATTGCCA CAATCTCTGG ATGGTTTTCC AGGGTTGTGT TATTACTGAG CTTCATCTTT 3060

CCAGAATGAG CAAAACACTG TCCAGTCTTT GTTACGATTT TGTAATAAAT GTGTACATTT 3120

TTTTTAAATT TTTGGACATC ACATGAATAA AGGTATGTAT GTACGAATGT GTATATATTA 3180

TATATATGAC ATCTATTTTG GAAAATGTTT GCCCTGCTGT ACCTCATTTT TAGGAGGTGT 3240

GCATGGATGC AATATATGAA AATGGGACAT TCTGGAACTG CTGGTCAGGG GACTTTGTCG 3300

CCCTGTGCAC TAAAAGGGCC AGATTTTCAG CAGCCAAGGA CATCCATACC CAAGTGAATG 3360

TGATGGGACT TAAAAGAAGT GAACTGAGAC AATTCACTCT GGCTGTTTGA ACAGCAGCGT 3420

TTCATAGGAA GAGAAAAAAA GATCAATCTT GTATTTTCTG ACCACATAAA GGCTTCTTCT 3480

CTTTGTAATA AAGTAGAAAA GCTCTCCTCA AAAAAAAAAA AAAAAAACTC GAG 3533

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1148 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

ACCCACGCGT CCGCAAATTA TACTTCCTCA TTCATATTAT GTTGATACAA AAGACCTTGG 60

CAGCCATTTC TCCCAGCAGT TTTAAAGGAT GAACATTGGA TTTCATGCCA TCCCATAGAA 120

AACCTGTTTT AAAATTTTAG GGATCTTTAC TTGGTCATAC ATGAAAAGTA CACTGCTTAG 180

AAATTATAGA CTATTATGAT CTGTCCACAG TGCCCATTGT CACTTCTTTG TCTCATTTCT 240

TCCCTTTGTT CCTTAGTCAT CCAAATAAGC CTGAAAACCA TAAGAGATAT TACTTTATTG 300

AATATGGTTG GCATTAAATT TAGCATTTCA TTATCTAACA AAATTAATAT AAATTCCAGG 360

ACATGGTAAA ATGTGTTTTA ATAACCCCCA GACCCAAATG AAAATTTCAA AGTCAATACC 420

AGCAGATTCA TGAAAGTAAA TTTAGTCCTA TAATTTTCAG CTTAATTATA AACAAAGGAA 480

CAAATAAGTG GAAGGGCAGC TATTACCATT CGCTTAGTCA AAACATTCGG TTACTGCCCT 540

TTAATACACT CCTATCATCA GCACTTCCAC CATGTATTAC AAGTCTTGAC CCATCCCTGT 600

CGTAACTCCA GTAAAAGTTA CTGTTACTAG AAAATTTTTA TCAATTAACT GACAAATAGT 660

TTCTTTTTAA AGTAGTTTCT TCCATCTTTA TTCTGACTAG CTTCCAAAAT GTGTTCCCTT 720

TTTGAATCGA GGTTTTTTTG TTTTGTTTTG TTTTCTGAAA AAATCATACA ACTTTGTGCT 780

TCTATTGCTT TTTTGTGTTT TGTTAAGCAT GTCCCTTGGC CCAAATGGAA GAGGAAATGT 840

TTAATTAATG CTTTTTAGTT TAAATAAATT GAATCATTTA TAATAATCAG TGTTAACAAT 900

TTAGTGACCC TTGGTAGGTT AAAGGTTGCA TTATTTATAC TTGAGATTTT TTTCCCCTAA 960

CTATTCTGTT TTTTGTACTT TAAAACTATG GGGGAAATAT CACTGGTCTG TCAAGAAACA 1020

GCAGTAATTA TTACTGAGTT AAATTGAAAA GTCCAGTGGA CCAGGCATTT CTTATATAAA 1080

TAAAATTGGT GGTACTAATG TGAAAAAAAA AAAAAAAAAA AACTCGAGGG GGGCCCGGTA 1140

CCCTATTA 1148

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GGCACGAGCT AGCTGCCGCC ACCCGAACAG CCTGTCCTGG TGCCCCGGCT CCCTGCCCCG 60

CGCCCAGTCA TGACCCTGCG CCCCTCACTC CTCCCGCTCC ATCTGCTGCT GCTGCTGCTG 120

CTCAGTGCGG CGGTGTGCCG GGCTGAGGCT GGGCTCGAAA CCGAAAGTCC CGTCCGGACC 180

CTCCAAGTGG AGACCCTGGT GGAGCCCCCA GAACCATGTG CCGAGCCCGC TGCTTTTGGA 240

GACACGCTTC ACATACACTA CACGGGAAGC TTGGTAGATG GACGTATTAT TGACACCTCC 300

CTGACCAGAG ACCCTCTGGT TATAGAACTT GGCCAAAAGC AGGTGATTCC AGGTCTGGAG 360

CAGAGTCTTC TCGACATGTG TGTGGGAGAG AAGCGAAGGG CAATCATTCC TTCTCACTTG 420

GCCTATGGAA AACGGGGATT TCCACCATCT GTCCCAGCGG ATGCAGTGGT GCAGTATGAC 480

GTGGAGCTGA TTGCACTAAT CCGAGCCAAC TACTGGCTAA AGCTGGTGAA GGGCATTTTG 540

CCTCTGGTAG GGATGGCCAT GGTGCCAGCC CTCCTGGGCC TCATTGGGTA TCACCTATAC 600

AGAAAGGCCA ATAGACCCAA AGTCTCCAAA AAGAAGCTCA AGGAAGAGAA ACGAAACAAG 660

AGCAAAAAGA AATAATAAAT AATAAATTTT AAAAAAAAAA AAAAAAAAAA AAAAAAA 717

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1099 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GGCACGAGCC GATGTGGACA TCATCCTGTC TATCCCCATG TTCCTGCGCC TGTACCTGAT 60

CGCCCGAGTC ATGCTGCTGC ACAGCAAGCT CTTCACCGAT GCCTCGTCCC GCAGCATCGG 120

GGCCCTCAAC AAGATCAACT TCAACACCCG CTTTGTCATG AAGACGCTCA TGACCATCTG 180

CCCTGGCACT GTGCTGCTCG TGTTCAGCAT CTCTCTGTGG ATCATTGCTG CCTGGACCGT 240

CCGTGTCTGT GAAAGTCCTG AATCACCAGC CCAGCCTTCT GGCTCATCAC TTCCTGCTTG 300

GTACCATGAC CAGCAGGACG TAACTAGTAA CTTTCTGGGT GCCATGTGGC TCATCTCCAT 360

CACATTCCTT TCCATTGGTT ATGGGGACAT GGTGCCCCAC ACATACTGTG GGAAAGGTGT 420

CTGTCTCCTC ACTGGCATCA TGGGTGCAGG CTGCACTGCC CTTGTGGTGG CCGTGGTGGC 480

CCGAAAGCTG GAACTCACCA AAGCGGAGAA GCACGTTCAT AACTTCATGA TGGACACTCA 540

GCTCACCAAG CGGATCAAGA ATGCTGCAGC CAATGTCCTT CGGGAAACAT GGTTAATCTA 600

TAAACACACA AAGCTGCTAA AGAAGATTGA CCATGCCAAA GTGAGGAAAC ACCAGAGGAA 660

GTTCCTCCCA AGCTATCCAC CAGTTTGAGG AGCGTCCCAG ATGGAACAGA GGAAAGCTGA 720

GTGACCAAGC CAACACTCTG GTGGACCTTT CCAAGATGCA GAATGTCATG TATGACTTAA 780

TCACAGAACT CAATGACCGG AGCGAAGACC TGGAGAAGCA GATTGGCAGC CTGGAGTCGA 840

AGCTGGAGCA TCTCACCGCC AGCTTCAACT CCCTGCCGCT GCTCATCGCC GACACCCTGC 900

GCCAGCAGCA GCAGCAGCTC CTGTCTGCCA TCATCGAGGC CCGGGGTGTC AGCGTGGCAG 960

TGGGCACCAC CCACACCCCA ATCTCCGATA GCCCCATTGG GGTCAGCTCC ACCTCCTTCC 1020

CGACCCCGTN CACAAGTTCA AGCAGTTGCT AAATAAATCT CCCCACTCCA GAAGCATTAA 1080

AAAAAAAAAA AAAAAAAAA 1099

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

AATTCGGCAG AGAGCCAACC GAGGGCGTTC CTGTCGGGGC TGCAGCGGCG GGAGGGAGCC 60

CAGTGGAGGC GCCCTCCCGA AGCGCCACTG CCCATGCTGA CCACCCAGCC CTCCGGCTGC 120

TGATGTCATG AGTAACACCA CTGTGCCCAA TGCCCCCCAG GCCAACAGCG ACTCCATGGT 180

GGGCTATGTG TTGGGGCCCT TCTTCCTCAT CACCCTGGTC GGGGTGGTGG TGGCTGTGGT 240

AATGTATGTA CAGAAGAAAA AGCGGGTGGA CCGGCTGCGC CATCACCTGC TCCCCATGTA 300

CAGCTATGAC CCAGCTGAGG AACTGCATGA GGCTGAGCAG GAGCTGCTCT CTGACATGGG 360

AGACCCCAAG GTGGTACATG GCTGGCAGAG TGGCTACCAG CACAAGCGGA TGCCACTGCT 420

GGATGTCAAG ACGTGACCTG ACCCCCTTGC CCCACCCTTC AGAGCCTGGG GTYCTGGACT 480

GCCTGGGGCC CTGCCATCTG CTTCCCCTGC TGTCACCTGG STCCCCCTGC TGGGTGCTGG 540

GTCTCCATTT CTCCCTCCAC CCACCCTCAG CAGCATCTGC TTCCCATGCC CTCACCATCA 600

CCTCACTGCC CCCAGGCCTT CTGCCCTTTG TGGGTGTTGA GCTCACCGCC CACCCACAGG 660

CACTCATGGG AAGAGGCTTT CCTTCTGGGA TGGCGGCGGC TGGTAGACAC CTTTGCTTTC 720

TCTAGCCCTC CTGGGCTGGG CTTGGGCACA AATCCCCAGG CAGGCTTTGG AGTTGTTTCC 780

ATGGTGATGG GGCCAGATGT ATAGTATTCA GTATATATTT TGTAAATAAA ATGTTTTGTG 840

GCTAAAAAAA AAAAAAAAAA ATCNAAGGGG GGGCCGGTAC CCAAATTCCC CCTATANTGA 900

ATTCGTATTA ACAATTCACT TGGGGCCGTC CTTTTAANAA C 941

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 756 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGCACGAGGA AGCTGGAGCG GGCCGGCGGT GCAGTCACGG GGGAGCGAGG CCTGCTGGGC 60

TTGGCAACGA GGGACTCGGC CTCGGAGGCG ACCCAGACCA CACAGACACT GGGTCAAGGA 120

GTAAGCAGAG GATAAACAAC TGGAAGGAGA GCAAGCACAA AGTCATCATG GCTTCAGCGT 180

CTGCTCGTGG AAACCAAGAT AAAGATGCCC ATTTTCCACC ACCAAGCAAG CAGAGCCTGT 240

TGTTTTGTCC AAAATCAAAA CTGCACATCC ACAGAGCAGA GATCTCAAAG ATTATGCGAG 300

AATGTCAGGA AGAAAGTTTC TGGAAGAGAG CTCTGCCTTT TTCTCTTGTA AGCATGCTTG 360

TCACCCAGGG ACTAGTCTAC CAAGGTTATT TGGCAGCTAA TTCTAGATTT GGATCATTGC 420

CCAAAGTTGC ACTTGCTGGT CTCTTGGGAT TTGGCCTTGG AAAGGTATCA TACATAGGAG 480

TATGCCAGAG TAAATTCCAT TTTTTTGAAG ATCAGCTCCG TGGGGCTGGT TTTGGTCCAC 540

AGCATAACAG GCACTGCCTC CTTACCTGTG AGGAATGCAA AATAAAGCAT GGATTAAGTG 600

AGAAGGGAGA CTCTCAGCCT TCAGCTTCCT AAATTCTGTG TCTGTGACTT TCGAAGTTTT 660

TTAAACCTCT GAATTTGTAC ACATTTAAAA TTTCAAGTGT ACTTTAAAAT AAAATACTTC 720

TAATGGAAAA AAAAAAAAAA AAAAAAAAAA ACTCGA 756

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2100 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

NCCAGAGGCA GAAAGTCCTG CTTCTGGGGC GTAACCTACA GGATATCCTT GGAACAGAAG 60
ATCTTATTGT GGAAGTRACT TCCAATGATG CTGTGAGATT TTATCCCTGG ACCATTGATA 120
ATAAATACTA TTCAGCAGAC ATCAATCTAT GTGTGGTGCC AAACAAATTT CTTGTTACTG 180
CAGAGATTGC AGAATCTGTC CAAGCATTTG TGGTTTACTT TGACAGCACA CAAAAATCGG 240
GCCTTGATAG TGTCTCCTCA TGGCTTCCAC TGGCAAAAGC ATGGTTACCY GAGGTGATGA 300
TCTTGGTCTG CGATAGAGTG TCTGAAGATG GTATAAACCG ACAAAAAGCT CAAGAATGGT 360
GCATCCAAAC ATGGCTTTGA ATTGGTAGAA CTTAGTCCAG AGGAGTTGCC TGAGGAGGAT 420
GATGACTTCC CAGAATCTAC AGGAGTAAAG CGAATTGTCC AAGCCCTGAA TGCCAATGTG 480
TGGTCCAATG TAGTGATGAA GAATGATAGG AACCAAGGCT TTAGCTTGCT GCAACTCATT 540
GACTGGAACA AACCATAGCA TTGGGTCAGC AGATCCCTGT CACCCAGAGC AACCCCATTT 600
GCCAGCAGCA GATAGTACTG AATCCCTCTC TGATCATCGG GGTGGTGCAT CTAACACAAC 660
AGATGCCCAG GTTGATAGCA TTGTGGATCC CATGTTAGAT CTGGATATTC AAGAATTAGC 720
CAGTCTTACC ACTGGAGGAG GAGATGTGGA GAATTTTGAA AGACTCTTTT CAAAGTTAAA 780
GGAAATGAAA GACAAGGCTG CGACGCTTCC TCATGAGCAA AGAAAAGTGC ATGCAGAAAA 840
GGTGGCCAAA GCATTCTGGA TGGCAATCGG GGAGACAGA GATGAAATTG AAGGCCTTTC 900
ATCTGATGAA GAGCACTGAA TTATTCATAC TAGGGTTTGA CCAACAAAGA TGCTAGCTGT 960
CTCTGAGATA CCTCTCTACT CAGCCCAGTC ATATTTTGCC AAAATTGCCC TTATCATGTT 1020
GGCTGCCTGA CTTGTTTATA GGGTCCCCTT AATTTTAGTT TTTAGTAGGA GGTTAAGGAG 1080
AAATCTTTTT TTTCCTCAGT ATATTGTAAG AGAGTGAGGA ATACAGTGAT AGTAATGAGT 1140
GAGGATTTCT TAAATRTACT TTTTTTTTGT TCTAGGAATG AGGGTAGGAT AAATCTCAGA 1200
GGTCTGTGTG ATTTACTCAA GTTGAAGACA ACCTCCAGGC CATTCCTGGT CAACCTTTTA 1260
AGTAGCATTT CCAGCATTCA CACTTGATAC TGCACATCAG GAGTTGTGTC ACCTTTCCTG 1320
GGTGATTTGG GTTTTCTCCA TTCAAGGAGC TTGTAGCTCT GAAGCTATGA TGCTTTTATT 1380
GGGAGGAAAG GAGGCAGCTG CAGAATTGAT GTGAGCTATG TGGGGCCGAA GTCTCAGCCC 1440
GCAGCTAAGT CTCTACCTAA GAAAATGCCT CTGGGCATTC TTTTGAAGTA TAGTGTCTGA 1500
GCTCATGCTA GAAAGAATCA AAAAGCCAGT GTGGATTTTT AGACTGTAAT AAATGAGGCA 1560

AAGGATTTCT ATTCCAGTGG GAAGRAAACC TCTCTACTGA GTTGTGGGGG ATATGTTGTA 1620

TGTTAGAGAG AACCTTAAGG AGTCCTTGTA TGGGCCATGG AGACAGTATG TGATAACATA 1680

CCGTGATTTT CATGAAGAAA TTCTTCTGTC TTAGAGTTCT CCCCTGCTGC TTGAGATGCC 1740

AGAGCTGTGT TGTTGCACAC CTGCAAAACA AGGCACATTT CCCCCTTTCT CTTTAAAGCC 1800

AAAGAGAGAT CACTGCCAAA GTGGGAGCAC TAAGGGGTGG GTGGGGAAGT GAAATGTTAG 1860

GCGATGAATT CCTGAGCACC TTGTTTTTCT TCCAAGGTTC GTAGCTCCTC TCTGCCCTTC 1920

CAAGCCTGTA ACCTCGGAGG ACTATCTTTT GTTCTTTATC CTTTGTCTTG TTTGAGTGGG 1980

TCAGCCCCAG AGGAACTGAT AAGCAAATGG CAAGTTTTTA AAGGAAGAGT GGAAAGTACT 2040

GCAAATAAAA ATCCTTATTT GTTTTTGTAG AAAAAAAAAA AAAAAAAAAA AAAAAAAAAG 2100

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

AAAAAAAAAA AAAGCCCACC TGAAAGCCTG TCTCTTTCCA CTTTGTTGGC CCTTCCAGTG 60

GGATTATCGA GCATGTTGTT TTTTCATAGT GCCTTTTTCC TTATTTCAAG GGTTGCTTCT 120

GAGTGGTGTT TTTTTTTTTT TTAATTTGTT TTGTTTTAAA ATAAGTTAAA GACAGTCCAG 180

AGCTTTTCAG CCAATTTGTC TCCTACTCTG TGTAAATATT TTTCCCTCCG GGCAGGGGAG 240

CCAGGGTAGA GCAAAGGAGA CAAGCAGGAG TGGAAGGTGA GGCGTTCTCC TGCTTGTACT 300

AAGCCAGGAG STTTAAGCTC CAGCTTTAAG GGTTGTGAGC CCCTTGGGGT TCAGGGAACT 360

GCTTGCCCAG GGTGCAGTGT GAGTGTGATG GGCCACCGGG GCAAGAGGGA AGGTGACCGC 420

CCAGCTCTCC CACATCCCAC TGGATCTGGC TTACAGGGGG GTCGGAAGCC TGTCCTCACC 480

GTCTCGGGGG TTGTGGCCCC CGCCCCCTCC CTATATGCAC CCCTGGAACC AGCAAGTCCC 540

AGACAAGGAG AGCGGAGGAG GAAGTCATGG GAACGCAGCC TCCAGTTGTA GCAGGTTTCA 600

CTATTCCTAT GCTGGGGTAC ACAGTGAGAG TACTCACTTT TCACTTGTCT TGCTCTTAGA 660

TTGGGCCATG GCTTTCATCC TGTGTCCCCT GACCTGTCCA GGTGAGTGTG AGGGCAGCAC 720

TGGGAAGCTG GAGTGCTGCT TGTGCCTCCC TTCCCAGTGG GCTGTGTTGA CTGCTGCTCC 780

CCACCCCTAC CGATGGTCCC AGGAAGCAGG GAGAGTTGGG GAAGGCAAGA TTGGAAAGAC 840

AGGAAGACCA AGGCCTCGGC AGAACTCTCT GTCTTCTCTC CACTTCTGGT CCCCTGTGGT 900

GATGTGCCTG TAATCTTTTT CTCCACCCAA ACCCCTTCCC ACGACAAAAA CAAGACTGCC 960

TCCCTCTCTT CCGGGAGCTG GTGACAGCCT TGGGCCTTTC AGTCCCAAAG CGGCCGATGG 1020

GAGTCTCCCT CCGACTCCAG ATATGAACAG GGCCCAGGCC TGGAGCGTTT GCTGTGCCAG 1080

GAGGCGGCAG CTCTTCTGGG CAGAGCCTGT CCCCGCCTTC CCTCACTCTT CCTCATCCTG 1140

CTTCTCTTTT CCTCGCAGAT GATAAAAGGA ATCTGGCATT CTACACCTGG ACCATTTGAT 1200

TGTTTTATTT TGGAATTGGT GTATATCATG AAGCCTTGCT GAACTAAGTT TTGTGTGTAT 1260

ATATTTAAAA AAAAAATCAG TGTTTAAATA AAGACCTATG TACTTAATCC TTTAACTCTG 1320

CGGATAGCAT TTGGTAGGTA GTGATTAACT GTGAATAATA AATACACAAT GAATTCTTMA 1380

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAACCCCGGG GGGGGCCCCG GGCCCCAATT 1440

CCCCCCAA 1448

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 456 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GGCACAGCAA ACTTGACGCC ATGAAGATCC CGGTCCTTCC TGCCGTGGTG CTCCTCTCCC 60

TCCTGGTGCT CCACTCTGCC CAGGGAGCCA CCCTGGGTGG TCCTGAGGAA GAAAGCACCA 120

TTGAGAATTA TGCGTCACGA CCCGAGGCCT TTAACACCCC GTTCCTGAAC ATCGACAAAT 180

TGCGATCTGC GTTTAAGGCT GATGAGTTCC TGAACTGGCA CGCCCTCTTT GAGTCTATCA 240

AAAGGAAACT TCCTTTCCTC AACTGGGATG CCTTTCCTAA GCTGAAAGGA CTGAGGAGCG 300

CAACTCCTGA TGCCCAGTGA CCATGACCTC CACTGGAAGA GGGGGCTAGC GTGAGCGCTG 360

ATTCTCAACC TACCATAACT CTTTCCTGCC TCAGGAACTC CAATAAAACA TTTTCCATCC 420

AAAAAAAAAA AAAAAAAAAC CCCNGGGGGG GCCCGG 456

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1326 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GGCACGAGTG CAGGCCCAGA GAGGACTCAT TGAAAGGACT GAAAGGGGAG GTGGCGTTTT 60
CTTCCTACCC AAACTTACCC CTGTGAGCTG GACAGCTTGG TAGCACCTGC CTGGACTTAG 120
ATGGTGGTAG CCAAGAAGAC TGACATTTTA GGGAACAGGA CGGGGAGGAG AAGGCTCTGG 180
CACACACACA TGTGTCCATA TGTCCTGCAA TGGTCTGGGG ACTATTGCTA GGCTAGGAGC 240
CCTAAGTGTC TTCTTCCTCA TGTCTMTTCT CCCCTGTSTC ATGGGCCCTA AGRTCTCTTT 300
CACTGGGCCT GCCTCAATGA ACGTGCTGCC CAGCTACCCC GAAACACGGC ANCTGCCGGC 360
TATCAATGCC CCAGCTGCAA TGGCCCATCT TCCCCCAACC AACCTGGCTG GGCCCGTGGG 420
CTCCGCACTG AGARARAAAS TTGGCACART CAACTGGGCC CGGGCAGGAC TGGGCCYCCC 480
TCTGATCGAT GAAGKTGGTG ARCCCAGAGC CCGAGCCCCT CAACACGTCT GACTTCTCTG 540
ACTGGTCTAG TTTTAATGCC AGCAGTACCC CTGGACCAGA GGAGGTAGAC AGCGCCTCTG 600
CTGCCCCAGC CTTCTACAGC CGAGCCCCCC GGCCCCCAGC TTCCCCAGGC CGGCCCGAGC 660
AGCACACAGT GATCCACATG GGCAATCCTG AGCCCTTGAC TCACGCCCCT AGGAAGGTGT 720
ATGATACGCG GGATGATGAC CGGACACCAG GCCTCCATGG AGACTGTGAC GATGACAAGT 780
ACCGACGTCG GCCGGCCTTG GGTTGGCTGG CCCGGCTGCT AAGGAGCCGG GCTGGGTCTC 840
GGAAGCGRCC GCTGACCCTG CTCCAGCGGG CGGGGCTGCT GCTACTCTTG GGACTGCTGG 900
GCTTCCTGGC CCTCCTTGCC CTCATGTCTC GCCTAGGCCG GGCCGCAGCT GACAGCGATC 960
CCAACCTGGA CCCACTCATG AACCCTCACA TCCGCGTGGG CCCCTCCTGA GCCCCCTTGC 1020
TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT AATGGGGAGG 1080
CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC 1140
AAAGCCAAGT CCACCAGAGT GGCTGCAGGC CAGGCCTGGA GTCCCCGTGG GTCAAGCATT 1200
TGTCTTGACT TGCTTTCCTC CCGGGTYTCC AGCCTCCGAC CCCTCGCCCC ATGAAGGAGC 1260
TGGCAGGTGG AAATAAACAA CAACTTTATT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1320
AAANAA 1326

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 710 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

```
GCGAAAGAGA AAAAGGCTGG AGCTCCCGCC CCCGGGGCTG TCAGATGGCT TGGGTTTCTG     60
CGACGCGATT GGCTCGCGGA GGGCAGAAAT TACTCAGCAA ACATGACTAT TATTAGCTGC    120
TTAGCAACAG CTCACCAAAG TAGAGAGACC ACCCAGGTAG GCAACCCAGT GTGTGCATCC    180
TCGGCTTCGG GGCAGCCTCT GAGAGCGCCA ACCTTCTCGC ATGCAATACT TCCATTAAGG    240
AATGCTCCCC CTCCTTTCTC TCTTATTCCT TTTCTTTTCA ACAGTGTCTT CTTTTTGTGG    300
GATGCCTTTG CGCGCACACA CGCGCGCGCA SGCACACACA CGAACATTTG CCTCGCGGTA    360
GACACGGGGG GAAATGTWAT ATTTTTTTAA GCGCTTAAAC AATTTCTGAA ATTCCTCAAA    420
GAAAAGCCTT TCAGARGCAC CTTGGCCTCA AGCTGCAACA AATACTGGGA RGTCCGGCTC    480
GCATTCCCAG GCCTGCACCA ATAATGACAG CGTGCTGGAT ARTGCGCCAG TGTGTGCCAG    540
ATTTTTTTTT CCTCTTCTCT TTTCTTTTAT AACTAAAGGG AAGACTTAGG CTCTTGCAGG    600
GAACAACGCC TCGCATTAAG ATAAACAGAA TGGAAAGTTA AAGAGGAAAG CAAGGACGTT    660
GGGAAAAGCC ATCTTTCTTA AAATCCGTCT GCCCCCCAGC CGCTTTCTCC               710
```

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1188 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

```
GATGGCTTTT ATATCTATTA TCGACCCACA GACAGTGACA ATGATAGTGA CTACAAGAAG     60
GATATGGTGG AAGGGGACAA GTACTGGCAC TCCATCAGCC ACCTGCAGCC AGAGACCTCC    120
TACGACATTA AGATGCAGTG CTTCAATGAA GGAGGGGAGA GCGAGTTCAG CAACGTGATG    180
ATCTGTGAGA CCAAAGCTCG GAAGTCTTCT GGCCAGCCTG GTCGACTGCC ACCCCCAACT    240
CTGGCCCCAC CACAGCCGCC CCTTCCTGAA ACCATAGAGC GGCCGGTGGG CACTGGGGCC    300
ATGGTGGCTC GCTCCAGCGA CCTGCCCTAT CTGATTGTCG GGGTCGTCCT GGGCTCCATC    360
GTTCTCATCA TCGTCACCTT CATCCCCTTC TGCTTGTGGA GGGCCTGGTC TAAGCAAAAA    420
CATACAACAG ACCTGGGTTT TCCTCGAAGT GCCCTTCCAC CCTCCTGCCC GTATACTATG    480
GTGCCATTGG GAGGACTCCC AGGCCACCAG GCAGTGGACA GCCCTACCTC AGTGGCATCA    540
GTGGACGGGC CTGTGCTAAT GGGATCCACA TGAATAGGGG CTGCCCCTCG GCTGCAGTGG    600
GCTACCCGGG CATGAAGCCC CAGCAGCACT GCCCAGGCGA GCTTCAGCAG CAGAGTGACA    660
```

CCAGCAGCCT GCTGAGGCAG ACCCATCTTG GCAATGGATA TGACCCCCAA AGTCACCAGA 720

TCACGAGGGG TCCCAAGTCT AGCCCGGACG AGGGCTCTTT CTTATACACA CTGCCCGACG 780

ACTCCACTCA CCAGCTGCTG CAGCCCCATC ACGACTGCTG CCAACGCCAG GAGCAGCCTG 840

CTGSTGTGGG CCAGTCAGGG GTGAGGAGAG CCCCCGACAG TCCTGTCCTG GAAGCAGTGT 900

GGGACCCTCC ATTTCACTCA GGGCCCCCAT GCTGCTTGGG CCTTGTGCCA GTTGAAGAGG 960

TGGACAGTCC TGACTCCTGC CAAGTGAGTG GAGGAGACTG GTGTCCCCAG CACCCCGTAG 1020

GGGCCTACGT AGGACAGGAA CCTGGAATGC AGCTCTCCCC GGGGCCACTG GTGCGTGTGT 1080

CTTTTGAAAC ACCACCTCTC ACAATTTAGG CAGAAGCTGA TATCCCAGAA AGACTATATA 1140

TTGTTTTTTT TTTAAAAAAA AAAAAAAAAA AWCYCGGGGG GGGGCCCC 1188

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 956 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GGCAGAGCAG TGAAAATGCA TCCTAAAAAT TCAATGTTTA TACCAGGCTC ATGACACTAA 60

GATGTGACAT CTGGACACGA GGGGTCAGCC ACGTGGATAC ATCCCTCCCA GATTGCATCT 120

CCAGGAATCA CTCTGCTAGC AGAATGGGCG CCCCATCCCT TACTATGCTG CTCCTCCTCA 180

AAGTGCAGCC CAGAAGGACC CAGGCCTTTG ATGCACATTG GGTGGGTCTC CCACTACTTT 240

AGTTGAAATG GGAGCATGCT GGAGTCGGCG TTCTGTTGCT TCTGGTGAGA AGGACATCCC 300

ATTGACCCCT GGCCACCAGG TCCAGTATTC CATCCTTCCT TCTGTCCCAG CCTATCGCCC 360

TCCCCACYAG GCCCACCCCC ACAACTTCTC CTCAAGGGAG GTTNTCCCGC AGCTGGAGGG 420

CTTGCACAGA CCAGCAGTCA CAGAAATCAT TCTTCCTGCT GTACTGGGCC TTAACTGCCT 480

GCAAATGTCC GAGCACTACT GCATAGGATG CCAGAGCCAC CGAAGATAAA CACAGCCAAG 540

TTTAATAATA ATAAAAGGAA AAATCTCAGC CTGCAGAACT CTGGTTTTGA CCCACCATCG 600

GCCAGATGCA CATCTTCAGG GCCTGTTGAG CACCTTCTGA AAAGCAGGGC TCGTAATAGA 660

CTCCAGCACA TTCCATCAGA GTCAGGAAAA CTGCGGTGAG TCCCAGAGAA TCTAGGGTGC 720

AGGGCAGGGA GCAGGAGTCA TAAGGAGTGA TAACCTAAAC TGTGTGTAGT CAGCGGGGAG 780

GGTCTTATGT TATCAGGTGA AATGAGAGCC AGTAAGTTAG TTGATCCTGT CACAGATATA 840

ACCCTGATAA CACCCCATAG ATACGCGACA CGTGTGTCCT GCCCCTGCTT TCCCCATCCA 900

ACATGGTTCT TCTGTTCCAC AGACATTAAA GGGGCTTTCT GCAATTACTT AAAAAA 956

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1603 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TCGACCCACG CGTCCGCTCT GCCAGGAATC TGGTCTTTCT GTAGACCCAA GTCAGAAAGA 60

ACCATTTGTG GAGTTAAATC GAATATTAGA RGCATTAAAR GTCAGAGTTC TGAGACCTGC 120

TCTGGAATGG GCAGTTTCAA ACCGAGAGAT GCTTATAGCC CAAAACAGCT CCTTGGAATT 180

TAAACTACAC AGACTGTATT TTATTAGCTT RTTAATGGGT GGAACACAAA TCAGCGAGAR 240

GCATTACAAT ATGCTAAAAA TTTTCAGCCA TTTGCCCTAA ATCATCAAAA AGACATTCAG 300

GTTTTGATGG GAAGCCTTGT GTACCTGAGA CAAGGGATTG AGAACTCACC ATATGTTCAC 360

CTACTTGATG CAAACCAGTG GGCTGATATC TGTGACATCT TTACACGGGA TGCTTGTGCC 420

CTCCTGGGGC TCTCCGTGGA GTCCCCTCTC AGTGTCAGTT TCTCAGCAGG TTGTGTGGCG 480

CTGCCAGCTT TAATTAACAT CAAAGCCGTG ATTGAACAGA GGCAGTGTAC TGGAGTTTGG 540

AACCAGAAAG ATGAATTACC TATTGAAGTG GACCTTGGTA AAAAGTGCTG GTATCACTCT 600

ATATTTGCCT GCCCCATTCT TCGTCAGCAA ACAACAGATA ACAATCCACC CATGAAATTG 660

GTCTGTGGTC ATATTATATC AAGAGATGCC CTGAATAAAA TGTTTAATGG TAGCAAATTA 720

AAATGTCCCT ACTGTCCAAT GGAACAAAGT CCAGGAGATG CCAAACAGAT ATTTTTCTGA 780

AGAGATAACT TTAGTTTGCA ATTTGTAAGT GAAACTGAAT CGTGGGTGCA TTTCAGAAGA 840

GAACGTTCCA TATAATGCAG CTAACCAAGG ACTCCTGTGT TTCTATAAGC TAATGCTCCA 900

GAAACTTTGC CAACCTGTTA GTGTACACAC ACTGAGGGGA GTGCTCCCGG TGAATATTAT 960

CATAGGGCTT TATTATATTC TTGGTCTTCA TTTCTGATCA AGTAAATACA CCAGCAGTTG 1020

TCATTCAATG CAGGTTTTTG TACTTAATTA TATGGTGATT TTTTTACTTT TTAAGAGCAG 1080

AAACGGAAAT TGACCTCCCC GCCATGTGTT TAATATTCCT CCTGCTTTTA CTTTTGTCAT 1140

TTTCTTGATA ATCGTAAGCC TTGAGAGTGT TTGTGAAAAA GTTTTATTTC CTGTTATGTA 1200

TACATAATTA AATGAAAATT CTTCAGAAAA AGTTTGATAA ATTGAATTGT GGTTATGAAA 1260

CTAATTTGCA TTTTTATTTG CTTAAGAAAG AAAGCTGTGA TAGATTCCAG ATATGCTTTT 1320

TGATGTTTTC CTCTGCTCCA GCTCCAAGAA GTCAGCACAC CTGCATTTTA GCTCTGCATG 1380

CAGCCCCAGC AGGCTGCGTG TTTAAGAATT TCATTGTTTA ACTGGCTGGT GTGAGAAGTC 1440

TTCCGTTAGC ATAGAGTGGA AGGAGTACTA TTGTTTGGTT GGGTTTTTGT TTGTTTGTTT 1500

TTTGTTTTTG CTTTTATTGC CAAGAGGTGC TTGTTTTAAA AGTATGTTTA ATAAAATGAA 1560

ATTCTAAAGT TAARAAGTGT TCTTAAAGTT GATATTTAAC TCT 1603

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1089 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GGCACGAGCT ACCTTTCTGC CTGCTTTGCT GGCTGCAACA GCACGAATCT CACGGGCTGT 60

GCGTGCCTCA CCACCGTCCC TGCTGAGAAC GCAACCGTGG TTCCTGGAAA ATGCCCCAGT 120

CCTGGGTGCC AAGAGGCCTT CCTCACTTTC CTCTGTGTGA TGTGTATCTG CAGCCTGATC 180

GGTGCCATGG CAAGACACCC TCAGTCATCA TCCTCATCAG GACAGTCAGC CCTGAACTCA 240

AGTCTTACGC TTTGGGAGTT CTTTTTCTCC TCCTTCGTTT GTTGGGCTTC ATCCCTCCAC 300

CCCTCATCTT CGGGGCTGGC ATCGACTCCA CCTGCCTGTT CTGGAGCACG TTCTGTGGGG 360

AGCAAGGCGC CTGCGTCCTC TACGACAATG TGGTCTACCG ATACCTGTAT GTCAGCATCG 420

CCATCGCGCT CAAATCCTTC GCCTTCATCC TGTACACCAC CACGTGGCAG TGCTGAGGAA 480

AAACTATAAA CGCTACATCA AAAACCACGA GGGCGGGCTG AGCACCAGTG AGTTCTTTGC 540

CTCTACTCTG ACCCTAGACA ACCTGGGGAG GGACCCTGTG CCCGCAAACC AGACACATAG 600

GACAAAGTTT ATCTATAACC TGGAAGACCA TGAGTGGTGT GAAAACATGG AGTCCGTTTT 660

ATAGTGACTA AAGGAGGGCT GAACTCTGTA TTAGTAATCC AAGGGTCATT TTTTTCTTAA 720

AAAAAGAAAA AAAGGTTCCA AAAAAAACCA AAACTCAGTA CACACACACA GGCACAGATG 780

CACACACACG CAGACAGACA CACCGACTTT GTCCTTTTTC TCAGCATCAG AGCCAGACAG 840

GATTCAGAAT AAGGAGAGAA TGACATCGTG CGGCAGGGTC CTGGAGGCCA CTCGCGCGGC 900

TGGGCCACAG AGTCTACTTT GAAGGCACCT CATGGTTTTC AGGATGCTGA CAGCTGCAAG 960

CAACAGGCAC TGCCAAATTC AGGGAACAGT GGTGGCCAGC TTGGAGGATG GACATTTCTG 1020

GATACACATA CACATACAAA ACAGAAAACA TTTTTTAAAA GAAGTTTCCT AAAATAAAAA 1080

AAAAAAAAA 1089

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 629 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGCTCAGTTC CCTTAGAAAT GAAATTTTAA ATGACACTAC CAGGTAAGCC ACTGAGACCA 60

GTGGAGGTGA TAGCTAAGAA CATAAGGAAT TAAGAATTTT TAATGGAGAA AGGAGGTAAT 120

GAATACCAGT TACATCCTAA GACTCACTGT AGTGGTGAGT GTTGTAATTT ATCTCGCTAT 180

CCATCCTCTT TTAAGTTTTT CCTTAGAAAG TCCTCTATTG GTACCTTGGA GGGACTGCTG 240

TCAAAATATA TGGAAAAGTG GGTCTGTGTG GTACAAGAGG TGGACTTTGC CACACATGGA 300

AGTTTGCTGC CAAGATCTTC ACTAATGAAA GAAATCACCA GTGAGCTGCA CAGATTAGCC 360

AAATACTGAG CTCATTAGAA CTACTAAGGC CTGGACATTT CTGCCTAATC CAGGACTCCT 420

GTAATTATCA GTCTTTGCTT TGGAGCTTCC CATTGTGTAG CTGARAATTT GTCATATCTG 480

CATTATAATC TAAGGCTCCA CATACTTAAT CCTGCTTCTC CCCCTTTTTC TTTCCCTTTC 540

CCAGCGGTCA GCTCTGCTGC ATAGTCTGAA GACTTTCCCT GCCCAATCCT GATAAAATTC 600

TTGCACTCGT AACCCCATCT CAGTGTCTG 629

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1964 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

AAGAAGACAT GGAAATTGCT GAAGGATGTT TCAGGCATAT TAAGAAAATC TTTACGCAGC 60

TTGAGGAATT CAGAGCCTCT GAATTGCTTC GAAGTGGACT GGACAGATCT AAATACCTTT 120

TAGTGAAAGA AGCCAAAATT ATTGCTATGA CCTGTACTCA TGCTGCCTTA AAACGACATG 180

ACTTGGTCAA GCTAGGTTTC AAGTATGACA ACATTTTGAT GGAAGAGGCT GCTCAGATTC 240

TGGAGATAGA AACTTTTATC CCTCTTCTTC TACAGAATCC TCAGGATGGA TTTAGCCGAC 300

TAAAACGATG GATTATGATT GGCGATCATC ACCAGTTACC TCCAGTTATT AANGAACATG 360

GCCTTTCAAA AGTACTCAAA CATGGAGCAG TCTCTCTTCA CTCGCTTTGT TCGCGTTGGA 420

GTTCCGACTG TTGACCTTGA TGCTCAAGGG AGAGCCAGAG CAAGCTTGTG CAMCTNCTAC 480

AACTGGCGAT ACAAGAATCT AGGAAACTTA CCCCATGTGC AGCTCTTGCC AGAGTTTAGT 540

ACAGCAAATG CTGGCTTACT GTATGACTTC CAGCTCATTA ATGTTGAAGA TTTTCAAGGA 600

GTGGGAGAAT CTGAACCTAA TCCTTACTTC TATCAGAATC TTGGAGAGGC AGAATATGTA 660

GTAGCACTTT TTATGTACAT GTGTTTACTT GGTTACCCTG CTGACAAAAT CAGTATTCTA 720

ACAACATATA ATGGCCAAAA GCATCTTATT CGCGACATCA TCAATAGACG ATGTGGAAAC 780

AATCCATTGA TTGGAAGACC AAACAAGGTG ACAACTGTTG ATAGATTTCA AGGTCAACAG 840

AATGACTATA TTCTTCTTTC TCTGGTACGA ACCAGGGCAG TGGGCCATCT GAGGGATGTC 900

CGTCGCTTGG TAGTGGCCAT GTCTAGAGCC AGACTTGGAC TTTATATCTT CGCCAGAGTA 960

TCCCTCTTCC AAAACTGTTT TGAACTGACT CCAGCTTTCA GTCAGCTCAC AGCTCGCCCC 1020

CTTCATTTGC ATATAATTCC AACAGAACCT TTCCCAACTA CTAGAAAGAA TGGAGAGAGA 1080

CCATCTCATG AAGTACAAAT AATAAAAAAT ATGCCCCAGA TGGCAAACTT TGTATACAAC 1140

ATGTACATGC ATTTGATACA GACTACACAT CATTATCATC AGACTTTATT ACAACTACCA 1200

CCTGCTATGG TAGAAGAGGG TGAGGAAGTT CAAAATCAAG AAACAGAATT GGAAACAGAA 1260

GAAGAGGCCA TGACTGTTCA AGCTGACATC ATACCCAGTC CAACAGACAC CAGCTGCCGT 1320

CAAGAAACTC CAGCCTTTCA AACTGACACC ACCCCCAGTG AGACAGGAGC CACTTCCACT 1380

CCAGAAGCCA TCCCTGCTTT ATCTGAGACC ACCCCTACTG TGGTAGGAGC TGTATCTGCA 1440

CCGGCAGAAG CTAACACACC TCAGGATGCC ACATCTGCCC CAGAAGAGAC CAAGTAGCCA 1500

AACTGTAGTC CTTCTAAAGG AGGACATGGC AGTCAAAAAG TCTGAGTAAA GCTGTTTTTT 1560

GTATTTTATA TTTGCTTCTG CCATTTTACT GTCACTAATT AATGTTTAGT TCTTATATTT 1620

GTTAACTGAT TTCGGTGTCT TGAATATATT TTTTTAAATT ATGTGTATGA ACAATTCTAG 1680

TTTCATTTGT TCAATCAGAA GAGCAAATAA CCATTCCTTT CATGTTTTGA TCACTGAGTG 1740

TGTCTGTAAT CATACCTACA TTAAAATCAT TTTCTATGAA TATATAATAT ATACTTCACA 1800

TTTTTAGTGA ACTTCTCTAA AGAAGAGGAC AGAATATACT GGACTTAACC ACGAATACCC 1860

TTGAGTGTCC AAATTGGGAA GGAACTKGTT TCTTCYGTTA TACTAYCAAA TGCTTAAATT 1920

CKGTTTCCTT TTTTCTTACC TTTGTTTGCT GTCTTTATGT AAAG 1964

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

```
CGTGTCCGCG CGCCTGGGAG ACGCTGCCTC GGCCCGGACG CGCCCGCGCC CCCGCGGCTG      60
GAGGGTGGTC GCCACTGGGA CACTGTGAAC CAGGAGTRAG TCGGAGCTGC CGCGCTGCCC     120
AGGCCATGGA CTGTGAGGTC AACAACGGTT CCAGCCTCAG GGATGAGTGC ATCACAAACC     180
TACTGGTGTT TGGCTTCCTC CAAAGCTGTT CTGACAACAG CTTCCGCAGA GAGCTGGACG     240
CACTGGGCCA CGAGCTGCCA GTGCTGGCTC CCCAGTGGGA GGGCTACGAT GAGCTGCAGA     300
CTGATGGCAA CCGCAGCAGC CACTCCCGCT TGGGAAGAAT AGAGGCAGAT TCTGAAAGTC     360
AAGAAGACAT CATCCGGAAT ATTGCCAGGC ACCTCGCCCA GGTCGGGGAC AGCATGGACC     420
GTAGCATCCC TCCGGGCCTG GTGAACGGCC TGGCCCTGCA GCTCAGGAAC ACCAGCCGGT     480
CGGAGGAGGA CCGGAACAGG GACCTGGCCA CTGCCCTGGA GCAGCTGCTG CAGGCCTACC     540
CTAGAGACAT GGAGAAGGAG AAGACCATGC TGGTGCTGGC CCTGCTGCTG GCCAAGAAGG     600
TGGCCAGTCA CACGCCGTCC TTGCTCCGTG ATGTCTTTCA CACAACAGTG AATTTTATTA     660
ACCAGAACCT ACGCACCTAC GTGAGGAGCT TAGCCAGAAA TGGGATGGAC TGAACGGACA     720
GTTCCAGAAG TGTGACTGGC TAAAGCTCGA TGTGGTCACA GCTGTATAGC TGCTTCCAGT     780
GTAGACGGAG CCCTGGCATG TCAACAGCGT TCCTAGAGAA GACAGGCTGG AAGATAGCTG     840
TGACTTCTAT TTTAAAGACA ATGTTAAACT TATAACCCAC TTTAAAATAT CTACATTAAT     900
ATACTTGAAT GAAAATGTCC ATTTACACGT ATTTGAATGG CCTTCATATC ATCCACACAT     960
GAATCTGCAC ATCTGTAAAT CTACACACGG TGCCTTTATT TCCACTGTGC AGGTTCCCAC    1020
TTAAAAATTA AATTGGAAAG CAGGTTTCAA GGAAGTAGAA ACAAAATACA ATTTTTTTGG    1080
TAAAAAAAAA TTACTGTTTA TTAAAGTACA ACCATAGAGG ATGGTCTTAC AGCAGGCAGT    1140
ATCCTGTTTG AGGAAAGCAA GAATCAGAGA AGGAACATAC CCCTTACAAA TGAAAAATTC    1200
CACTCAAAAT AGGGACTATC YATCTTAATA CTAAGGAACC AACAATCTTC CTGTTTAAAA    1260
AACCACATGG CACAGAGATT CNGAACTAAA GTGCTGCACT CAAATGATGG GAAGTCCCGG    1320
CCCCAGTACA CCAGGGGCTT TGGACTTTTT TCAACTTCGT TTCCTTTTGT TTGGANTCCA    1380
AAAGAACCAC TTTGTGGTTC TTAAAAGGGT GTGAAGGTGA TTTAAGGGGC CCAGGTCAGC    1440
CACTGGTTGG TTTACAAAAT CNGGGTAACT AACTGCATAC AACTTTTTCC CNTTTCCATG    1500
NCATCAGGAC TTTGCTAAAG AC                                             1522
```

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 875 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

```
TGGGATTTCC CTTTATCATG GAGGCCTTGT CCCACTTCCT CTATGTCCCT TTCCTTGGTG     60
TCTGTGTCTG TGGGGCCATC TACACTGGCC TGTTCCTTCC TGAGACCAAA GGCAAGACCT    120
TCCAAGAGAT CTCCGAGGAA TTACACAGAC TCAACTTCCC CAGGCGGGCC CAGGGCCCCA    180
CGTGGAGGAG CCTGGAGGTT ATCCAGTCAA CAGAACTCTA GTCCCAAAGG GGTGGCCGTA    240
GCCAAAGCCA GCTACCGTCC TGTCCTCTGC TTCCTGCCAG GGCCCTGGTC CTCAMTYCCT    300
YCTGCATTCC TCATTTAAGG AGTGTTTATT GAGCACCCTT TGTGTGCAGA CATGGCTCCA    360
GGTGCTTAGC AATCAWTGGT GAGCGTGGTA TCCAGGCTAA AGGTAATTAA CTGACAGRAA    420
ATCAGTAACA ACATAATTAC AGGYTGGTTG TGGCAGYTCA TGACTGTAAT CCCAGCACTT    480
TTGGGAGCCA AGGTGGGARG ATCAATTGAG GCCAGAGTTT GAAAMCAGCT AGGTAACATA    540
GTGAGACCCC CTATCTCTAC AAAAAATTTT AAACATTAGC TGGGCATGGT GGTATGTGCT    600
AACAGCTCTA GCTACTCAGG AGGCTGAGGC AGCAGGATCA CTTGAGTCCA AGAGTTCAAG    660
GTAGCAGTAA GCTACAATCA CACCACTGCA TGCCAGACTG GGTGACAGAG GGAGACTTCA    720
TCTCTTTAAA ACATAATAAT AATAATTACA GACTCAGGAA ATGCAGTGAA AGAAAAATAC    780
AGGTTGGCCA GGTGAGGTGG CTGATGCCTG TAATCCCAGC ACTTTGGGAG GCCAAGATGG    840
GAAGATTGCT TTGAGACCAG AAGTTTGAGA CCAGC                               875
```

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 843 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

```
CCCACGCGGT CCGNATCGTC CTTCCCTCAC TTCAGAGGGT GGCCAGAGCT GAATACCCAG     60
AGAGGGACAA GTAAGGGTCC AGTTCCAAAA CATCATGAGG ATGTATCATC CCACGTGTCT    120
```

CACCTGACAG TTACAGAGGA AACCCGCACC CAGAATGCAC GTGCTGTCTT ATGGGAACAC 180

TCAGCGCAGA GTGCTCAGGT CCGGCCACAC TCGGGCTGTG CTTGGTCGTG CCATGGAATT 240

CCTCAGGACT TTCTCAGCCT CCCTAATGGC AGAAGCCCCT TTACAGCAAG ACATTTACCG 300

TTTGTCTGAA AATAGCCGAA CTGAGCTTTT CTTCAGGCTA TATGAGAAGT CTCTAGACAG 360

TGGGCACCGT CAGAAAGCCC AGAGCCTTGT GATAGCTCCC ACCCTGCCTG GCTCAGATCT 420

TCCCATTTTT TTTCCTCTGG CACTAACCTC ACCTTTTGTT TTTTTGTGTT TGTGTTTGTT 480

TTTGTTTTTG CAGAGTTGGA TTACAGAAAC TCCTATGAAA TTGAATATAT GGAGAAAATT 540

GGCTCCTCCT TACCTGTAAG TTCGTCTGCC TCGGGCCACT TAGGGGACTC GCTTTCCTGC 600

CTTCAGGGGC CTCCTCCCCT GTGCAGAGTG TCTCTGGGAG CTCAGACCCC AAATCGAGTG 660

TTTTCTGTGT ACACAGCTTC CCGGGTGCAC AGCAATGATG GACTGGGGCT GGGGGGTTGA 720

GGTTTGTACT CAATCCACTT CGTTTGACAT TTTCAGGGAG AAAATGATAG AATACAATTA 780

GACGTCCTGC AGAATTACTT TCCTAGACTG AGAAAGAGCT AGAGATTTCT TTAAAAAAAA 840

AAA 843

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 489 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

CTCTTAGGCT TTGAAGCATT TTTGTCTGTG CTCCCTGATC TTCAGGTCAC CACCATGAAG 60

TTCTTAGCAG TCCTGGTACT CTTGGGAGTT TCCATCTTTC TGGTCTCTGC CCAGAATCCG 120

ACAACAGCTG CTCCAGCTGA CACGTATCCA GCTACTGGTC CTGCTGATGA TGAAGCCCCT 180

GATGCTGAAA CCACTGCTGC TGCAACCACT GCGACCACTG CTGCTCCTAC CACTGCAACC 240

ACCGCTGCTT CTACCACTGC TCGTAAAGAC ATTCCAGTTT TACCCAAATG GGTTGGGGAT 300

CTCCCGAATG GTAGAGTGTG TCCCTGAGAT GGAATCAGCT TGAGTCTTCT GCAATTGGTC 360

ACAACTATTC ATGCTTCCTG TGATTTCATC CAACTACTTA CCTTGCCTAC GATATCCCCT 420

TTATCTCTAA TCAGTTTATT TTCTTTCAAA TAAAAAATAA CTATGAGCAA CAAAAAAAAA 480

AAAAAAAAA 489

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GAAGCAGTGT GTATCTATGA TTATATCTCT GTTCATCTAT ATATTTTTGA CATGTAGCAA 60

CACCTCTCCA TCTTATCAAG GAACTCAACT CGGTCTGGGT CTCCCCAGTG CCCAGTGGTG 120

GCCTTTGACA GGTAGGAGGA TGCAGTGCTG CAGGCTATTT TGTTTTTTGT TACAAAACTG 180

TCTTTTCCCT TTTCCCCTCC ACCTGATTCA GCATGATCCC TGTGAGCTGG TTCTCACAAT 240

CTCCTGGGAC TGGGCTGAGG CAGGGGCTTC GCTCTATTCT CCCTAACCAT ACTGTCTTCC 300

TTTCCCCTTG CCACTTAGCA GTTATCCCCC CAGCTATGCC TTCTCCCTCC CTCCCTTGCC 360

CTGGCATATA TTGTGCCTTA TTTATGCTGC AAATATAACA TTAAACTATC AAGTGAAAAA 420

AAAAAAAAAA AAAACTCCAA GGGGGGGCCG GTACCCAATT CCCCCTATAN TGAGTCNTAT 480

TACAATTCAC TGGGCCGTCG TTTTACAACG TCGTGAATGG GAAAACCTGG GCGT 534

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1374 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GGCACGAGTC CGGGATGAGC TCAGCCGCGG CCGACCACTG GGCGTGGTTG CTGGTGCTCA 60

GCTTCGTGTT TGGATGCAAT GTTCTTAGGA TCCTCCTCCC GTCCTTCTCA TCCTTCATGT 120

CCAGGGTGCT GCAGAAGGAC GCGGAGCAGG AGTCACAGAT GAGAGCGGAG ATCCAGGACA 180

TGAAGCAGGA GCTCTCCACA GTCAACATGA TGGACGAGTT TGCCAGATAT GCCAGGCTGG 240

AAAGAAAGAT CAACAAGATG ACGGATAAGC TCAAAACCCA TGTGAAAGCT CGGACAGCTC 300

AATTAGCCAA GATAAAATGG GTGATAAGTG TCGCTTTTCTA CGTATTGCAG GCTGCCCTGA 360

TGATCTCACT CATTTGGAAG TATTATTCTG TCCCTGTGGC TGTCGTGCCG AGTAAATGGA 420

TAACCCCTCT AGACCGCCTG GTAGCCTTTC CTACTAGAGT AGCAGGTGGT GTTGGAATTA 480

CCTGTTGGAT TTTAGTCTGT AACAAAGTTG TCGCTATTGT GCTTCATCCG TTCAGCTGAA 540

CAGGAGGATG GATACAGCCG CGAGGCTAAA AAACGGATTT CCTCTTCCTA GCTTAAAATC 600

TGATTTACAC TGTTTTGTTT TTTAAGAAAC AAAAGTGCAT AGTTTAGATT TTTTTTTTTG 660

TTGAATATGT TTGTTCTTGG ACTTTATGAG AGAGTCTTAT AAGAATCACG ATTTTCTACA 720

CCTGTCATTG AGCCAAGAAA GTCCAGTTTA TGACACGTAT GTACTAGTGA ACACCGTCCT 780

CGATCTGTAC GAAATGTGAA ATGTTTAGGG ACATCTCCAT GCTGTCACTT GTGATTTGCC 840

CTCTTATGTA TTTTGGTCAT ATTGCCAACT GGAAAGTCAA AATTTTCTAA CAACTTTAAG 900

TAAGTTCTTT GAAGACTTAG TGCTGTTTTT AATCCAGTTT AGAAAGTAAC TTAATTTTAA 960

TACCACTACT AAAAATTCGA AAATTTCTTC TTTAATCACA TTCAATATGG TTAAAAGAAC 1020

AACACTAATT GACATTGCGT GGGCTTTTTC TCCCTTTGTT TAAAATGTCA TTTGTTGAGC 1080

AAGAGTTGTA TAGTATTATC TACTTACTTG AGGCTGTTAA TTTTTCATTA CAGTGTTTTG 1140

TAAATGTATC CACGAGACCA TGATGCATTG TTTTGTGCTC AACTTGTGTT TTGTATTTAA 1200

AGCATTTTGA ATGAAGTGTA TTTTATAAGC ATTTAATATT TATGCTCTTT AGAATGGAAC 1260

ACAGAAAACA AACCTTATAA GTCCTGATTA ATCTGAACCA ATAACCTGTG TGGCCTACAA 1320

AGTATAATTC TATTAAATGT TCCTTAAAAC AAAAAAAAAA AAAAAAAAAA AAAA 1374

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 596 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

GAATTCGNCA CGAGATTACT TGGACATGAA AGAACTCAGG TTCAAGTTTA TTCATTTACT 60

AAGTTAGTTA AATCATGTGC CTTCCATGAG CCTTCATTTG GTAACTTGGA AAATGGAAAT 120

AATAACACTA GTCATATATA TTCTACACTG CTACCATATG GACCAAAGGG ATTATAGATT 180

ACAATCACCA TCATTCCTGC TGACAGGTAT ATAGAAAACA ATTTCATTGA AGAAAAGTCC 240

TTACATTTAT CCTTTTCCTA ATATCTGCAT GGGTAAACTA ATAAATATAG TCATTAGAAA 300

ACCCTTATTA TTATTATTAG TTCAATGTGA GAACTGCTGC AGAAAAAATA TGCTTTATAA 360

TATTTTCTTG AATATACATA ATATTCATAA ATTTTCAAAT CATTGAAAAT TACCTTAAAA 420

TTGGAAAAAA TGTGCATTTC TACTCATATA ACAGTATAAA ATTCCTATGT CAATCTCTTT 480

TTTTTTTTTT TGTTTTGAGT TGGAGTCTCG CTCTGTCGCC CAGGCTGGGC AACAGAGCAG 540

GACCCTGTCT TAATTAAAAA AAAAAAAAAA AAACTCGAGG GGGGCCCGGT ACCCTA 596

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 851 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

```
CACATGAAGA CACACAGTGG TGAGAAGCCC TTCCGCTGCG CCCGCTGTCC TTATGCCTCT      60
CCTCATCTGG ATAACCTGAA ACGGCACCAG CGCGTCCATA CAGGAGAGAA GCCCTACAAG     120
TGCCCCCTCT GCCCTTATGC CTGTGGCAAT CTGGCCAACC TCAAGCGTCA TGGTCGCATC     180
CACTCTGGTG ACAAACCTTT TCGGTGTAGC CTTTGCAACT ACAGCTGCAA CCAGAGCATG     240
AACCTCAAAC GTCACATGCT GCGGCACACA GGCGAGAAGC CTTCCGCTGT GCCACCTGCG     300
CCTATACCAC GGGCCACTGG GACAACTACA AGCGCCACCA GAAGGTGCAT GGCCACGGTG     360
GGGCAGGAGG GCCTGGTCTC TCTGCCTCTG AGGGCTGGGC CCCACCTCAT AGCCCACCCT     420
CTGTTTTGAG CTCTCGGGGC CCACCAGCCC TGGGGACTGC TGGCAGCCGG GCTGTCCACA     480
CAGACTCATC CTGAACTAGG TCCTTCTTCC CCATGTTTTA TACAGACGGA CCAGAAGCCA     540
CCTTTTTCTC CCCCGCTGGC CAGGGGCTCC ACACAGACTA ACGTAGGCAC TATAAGGACC     600
AGCCCAACCC CATGGGCGGG GGGGCCCATA TGGACCAGGG GACCTTGCCT TGACTGAGGC     660
ACTTCACGAG CTCAGTGAGA AGGGCCCTGT ATTCACCTCC ACTGCCCCCA GGGGCTGTGG     720
ACAAACCGGC TGGGGGACTG CCCAGCCTCC CACCTGTTTA TTTAACTTAT TTCAGTGCTT     780
TATAATAAAG GAAACACTAA CAAAGCCATG TCTATGCTGA ATTGGCAATG GCAGGCAATT     840
TGGCCTTACC C                                                          851
```

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

```
GTGAAATGAA AACAGTCTTT TTATAGCCTT TAGCTTGTGA GTTTGGAAGT TTGGGGGGTC      60
TTATGTTTGT TTTGCCTCTT CTGTTTCTTG GAGGAGAGTT GAGGCTTTTC TTAGGTGCAT     120
```

ACACAGACCC AGGTGAACAC GCTGACTGTG AACCTGCCCT GTATCCGGAG CTGTGCTGGG 180
CACTGAGGGG ATGCAACAAA ATTAGGAGAG GWTCCTTGCT CCCAACGTCT ACTTCTCCTA 240
CCTCAACAGG GGTCCAGGGT GCAGTGAACT CAGTTCTTGG CCCTTGGGTG AGGATTCATG 300
GATGAATGAA AGCTAGACCT GATGGGGAGG CATTATGACT AAATAGGCCC AGCCTCCTTC 360
CCTTCCAGCT CTGTCCTAGG AGCATAGGCG GGAAATCTGA GTAGAGTCTG ACTGCAGTTT 420
TTGCTTATGA TTTGTAAAAG CCGTCATGGG GTCAATAAGA AAATAGGGGT GATGGAGGGG 480
GAGAAGCCCA GGACTGGGAG AATCGCACGT GCCCCAGGGG TTTTCACCAA GGATTTTCAA 540
GACAAACTGG AGTAAGAATT AAAGCCCCAG AGGATTTAAT TATCCTGGTT TGCAAAAGAG 600
CCTCCCATGC CAGTACCGCC CAGCCTTGGA GGCCGGAATG CTCATGGCCC CTGTGGTCTG 660
CTTGTCCTTC AGCCCATGCC CAGCAGATAC CTCTCTGACT GGAGACGGGC TCAAAGCTGG 720
ATTAGAAAGG GGAGMGGCAC TTGTGACTTT GTTTGACTCT GTGACTCACT TCCTCGCTCA 780
CACCTTGTTT GAACTACTGG ACTTTCAACT GGCTTTCCTT AGGTCAGGCA AGCAGACAGC 840
TCCCCACTGA AGAGGTCTGT ACAGTGACAA CCCGGGCCGG CAGCAAGGAC ACAGATGCAG 900
CCACAGTAAG GCTCCATCAG GACTGGGTCA GTGATGGCAA CAGGATGGCC AAGGATGGCT 960
CTAGAACAYT CTGTCCATGC GTCACTCCCC CCAGTTTTRT TTTTAGCTTT GGCTTCAGGG 1020
AGTGACAGCC ATCACAAATA GCCACATTCT GCTCTACTCT CCAACATACC AGATTSTACA 1080
CTGTTGTTAT TTCATGAGAC GTGAATGTTG CAGAGAGTGG GGGGATTCTG GTTGTTAAGG 1140
AACTTACACT GGGGAGCTTT ACTCTTCCGT GTCAACAATG TGACTACATG TTCTCCAGAT 1200
TAGCCACACA TGCAAACATC AGTGTCCTTC TAGCTTTANC CGAGAAAGAA ACCAGTCCCA 1260
GGGAATGAAT GGTGGTCTCC CCACTCCCGG CAGCACTTTA GGCAGCCCAT AAGCTATGCG 1320
AGAATGTGAA CGCTCACCTT GCTCCGTCAC GGTTCTGACC TACCACATAA ACAGGAAGAA 1380
GCCAGTGACC GGAACAGCTC TAGGAATAAC AAGTCAGAAT AGAAGTGTCC TTTATATTAC 1440
CAGAAAATAT GGGCTTGGCC TAAGTCGCTG TCTCCTAACC TGCCGGGGTC ATTCCCCACC 1500
AAACACCCCA TACTAAGGAG CCATGAGCCA CCTGGACATT CACCTTTTCT TTGACCATCT 1560
GGAGTCTGGG GCAACTTAAG GAAGGCNCCA CACAGTGGTG CAGGCACATT TCCAAGCGTA 1620
GGTGTCCCTG GCTTTTGTGG CCAAAGCTAG TGTTATGGTC AACAACAGGC CAGGGTCTGT 1680
GGGGCACTGA CCTTGAAAGT GGCAAAATGG AGGTTTCACA GGCTGTGCGG GAGCAGGACG 1740
GCTTGCTTCA TCTAACAATC TCAGTTTCCT TTAAAAAAAG AAAGAAAGGA AAAGATTTCA 1800
TAAGCAGGTG TCAGTGGACA GTTTAAGYAC TTAACCATTT CTCTTTCTTC TTATGGATGT 1860
GAACTGTGCT GTGGATAAAT CATTTGTATT TCTTGAATGT TCTCTATGAC TAACAGTTAT 1920

TAAGTCGGTT GTGTATATGT GTAACTAATG TAACTGCCTT TTAAAATTTC ATTACAATAA 1980

AAATGACTTT GCTCTGAAMA AAAAAAAAAA AAAAACTCGA 2020

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2432 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

ATGAAGGGTC GTTGGTGGGA AAGATGGCGG CGACTCTGGG ACCCCTTGGG TCGTGGCAGC 60

AGTGGCGGCG ATGTTTGTCG GCTCGGGATG GGTCCAGGAT GTTACTCCTT CTTCTTTTGT 120

TGGGGTCTGG GCAGGGGCCA CAGCAAGTCG GGGCGGGTCA AACGTTCGAG TACTTGAAAC 180

GGGAGCACTC GCTGTCGAAG CCCTACCAGG GTGTGGGCAC AGGCAGTTCC TCACTGTGGA 240

ATCTGATGGG CAATGCCATG GTGATGACCC AGTATATCCG CCTTACCCCA GATATGCAAA 300

GTAAACAGGG TGCCTTGTGG AACCGGGTGC CATGTTTCCT GAGAGACTGG GAGTTGCAGG 360

TGCACTTCAA AATCCATGGA CAAGGAAAGA AGAATCTGCA TGGGGATGGC TTGGCAATCT 420

GGTACACAAG GAATCGGATG CAGCCAGGGC CTGTGTTTGG AAACATGGAC AAATTTGTGG 480

GGCTGGGAGT ATTTGTAGAC ACCTACCCCA ATGAGGAGAA GCAGCAAGAG CGGGTATTCC 540

CCTACATCTC AGCCATGGTG AACAACGGCT CCCTCAGCTA TGATCATGAG CGGGATGGGC 600

GGCCTACAGA GCTGGGAGGC TGCACAGCCA TTGTCCGCAA TCTTCATTAC GACACCTTCC 660

TGGTGATTCG CTACGTCAAG AGGCATTTGA CGATAATGAT GGATATTGAT GGCAAGCATG 720

AGTGGAGGGA CTGCATTGAA GTGCCCGGAG TCCGCCTGCC CCGCGGCTAC TACTTCGGCA 780

CCTCCTCCAT CACTGGGGAT CTCTCAGATA ATCATGATGT CATTTCCTTG AAGTTGTTTG 840

AACTGACAGT GGAGAGAACC CCAGAAGAGG AAAAGCTCCA TCGAGATGTG TTCTTGCCCT 900

CAGTGGACAA TATGAAGCTG CCTGAGATGA CAGCTCCACT GCCGCCCCTG AGTGGCCTGG 960

CCCTCTTCCT CATCGTCTTT TTCTCCCTGG TGTTTTCTGT ATTTGCCATA GTCATTGGTA 1020

TCATACTCTA CAACAAATGG CAGGAACAGA GCCGAAAGCG CTTCTACTGA GCCCTCCTGC 1080

TGCCACCACT TTTGTGACTG TCACCCATGA GGTATGGAAG GAGCAGGCAC TGGCCTGAGC 1140

ATGCAGCCTG GAGAGTGTTC TTGTCTCTAG CAGCTGGTTG GGGACTATAT TCTGTCACTG 1200

GAGTTTTGAA TGCAGGGACC CCGCATTCCC ATGGTTGTGC ATGGGGACAT CTAACTCTGG 1260

TCTGGGAAGC CACCCACCCC AGGGCAATGC TGCTGTGATG TGCCTTTCCC TGCAGTCCTT 1320

CCATGTGGGA GCAGAGGTGT GAAGAGAATT TACGTGGTTG TGATGCCAAA ATCACAGAAC 1380

AGAATTTCAT AGCCCAGGCT GCCGTGTTGT TTGACTCAGA AGGCCCTTCT ACTTCAGTTT 1440

TGAATCCACA AAGAATTAAA AACTGGTAAC ACCACAGGCT TTCTGACCAT CCATTCGTTG 1500

GGTTTTGCAT TTGACCCAAC CCTCTGCCTA CCTGAGGAGC TTTCTTTGGA AACCAGGATG 1560

GAAACTTCTT CCCTGCCTTA CCTTCCTTTC ACTCCATTCA TTGTCCTCTC TGTGTGCAAC 1620

CTGAGCTGGG AAAGGCATTT GGATGCCTCT CTGTTGGGGC CTGGGGCTGC AGAACACACC 1680

TGCGTTTCAC TGGCCTTCAT TAGGTGGCCC TAGGGAGATG GCTTTCTGCT TTGGATCACT 1740

GTTCCCTAGC ATGGGTCTTG GGTCTATTGG CATGTCCATG GCCTTCCCAA TCAAGTCTCT 1800

TCAGGCCCTC AGTGAAGTTT GGCTAAAGGT TGGTGTAAAA ATCAAGAGAA GCCTGGAAGA 1860

CATCATGGAT GCCATGGATT AGCTGTGCAA CTGACCAGCT CCAGGTTTGA TCAAACCAAA 1920

AGCAACATTT GTCATGTGGT CTGACCATGT GGAGATGTTT CTGGACTTGC TAGAGCCTGC 1980

TTAGCTGCAT GTTTTGTAGT TACGATTTTT GGAATCCCAC TTTGAGTGCT GAAAGTGTAA 2040

GGAAGCTTTC TTCTTACACC TTGGGCTTGG ATATTGCCCA GAGAAGAAAT TTGGCTTTTT 2100

TTTTCTTAAT GGACAAGAGA CAGTTGCTGT TCTCATGTTC CAAGTCTGAG AGCAACAGAC 2160

CCTCATCATC TGTGCCTGGA AGAGTTCACT GTCATTGAGC AGCACAGCCT GAGTGCTGGC 2220

CTCTGTCAAC CCTTATTCCA CTGCCTTATT TGACAAGGGG TTACATGCTG CTCACCTTAC 2280

TGCCCTGGGA TTAAATCAGT TACAGGCCAG AGTCTCCTTG GAGGGCCTGG AACTCTGAGT 2340

CCTCCTATGA ACCTCTGTAG CCTAAATGAA ATTCTTAAAA TCACCGATGG AACCAAAAAA 2400

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA 2432

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2340 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GACGCTGGGG GCGGGTGGGG GCGCGGGGTA CCGGGCTGGA CGGCCGGCCG GCGCCCCCTC 60

ATTAGTATGC GGACGAAGCG GCGGGCTGCG CGGAGNGACG TCCCCTGCAG CCGCGGACCG 120

AGGCAGCGGC GGCACCTGCC GGCCGAGCAA TGCCAAGTGA GTACACCTAT GTRAAACTGA 180

GAAGTGATTG CTCGAGGCCT TCCCTGCAAT GGTACACCCG AGCTCAAAGC AAGATGAGAA 240

GGCCCAGCTT GTTATTAAAA GACATCCTCA AATGTACATT GCTTGTGTTT GGAGTGTGGA 300

TCCTTTATAT CCTCAAGTTA AATTATACTA CTGAAGAATG TGACATGAAA AAAATGCATT 360

ATGTGGACCC TGACCATGTA AAGAGAGCTC AGAAATATGC TCAGCAAGTC TTGCAGAAGG 420

AATGTCGTCC CAAGTTTGCC AAGACATCAA TGGCGCTGTT ATTTGAGCAC AGGTATAGCG 480

TGGACTTACT CCCTTTTGTG CAGAAGGSCC CCAAAGACAG TGAAGCTGAG TCCAAGTACG 540

ATCCTCCTTT TGGGTTCCGG AAGTTCTCCA GTAAAGTCCA GACCCTCTTG GAACTCTTGC 600

CAGAGCACGA CCTCCCTGAA CACTTGAAAG CCAAGACCTG TCGGCGCTGT GTGGTTATTG 660

GAAGCGGAGG AATACTGCAC GGATTAGAAC TGGGCCACAC CCTGAACCAG TTCGATGTTG 720

TGATAAGGTT AAACAGTGCA CCAGTTGAGG GATATTCAGA ACATGTTGGA AATAAAACTA 780

CTATAAGGAT GACTTATCCA GAGGGCGCAC CACTGTCTGA CCTTGAATAT TATTCCAATG 840

ACTTATTTGT TGCTGTTTTA TTTAAGAGTG TTGATTTCAA CTGGCTTCAA GCAATGGTAA 900

AAAAGGAAAC CCTGCCATTC TGGGTACGAC TCTTCTTTTG GAAGCAGGTG GCAGAAAAAA 960

TCCCACTGCA GCCAAAACAT TTCAGGATTT TGAATCCAGT TATCATCAAA GAGACTGCCT 1020

TTGRACATCC TTCAGTACTC AGAGCCTCAG TCAAGGTTCT GGGGGCCGAG ATAAGAACGT 1080

CCCCACAATC GGTGTCATTG CCGTTGTCTT AGCCACACAT CTGTGCGATG AAGTCAGTTT 1140

GGCGGGTTTT GGATATGACC TCAATCAACC CAGAACACCT TTGCACTACT TCGACAGTCA 1200

ATGCATGGCT GCTATGAACT TTCAGACCAT GCATAATGTG ACAACGGAAA CCAAGTTCCT 1260

CTTAAAGCTG GTCAAAGAGG GAGTGGTGAA AGATCTCAGT GGAGGCATTG ATCGTGAATT 1320

TTGAACACAG AAAACCTCAG TTGAAAATGC AACTCTAACT CTGAGAGCTG TTTTTGACAG 1380

CCTTCTTGAT GTATTTCTCC ATCCTGCAGA TACTTTGAAG TGCAGCTCAT GTTTTTAACT 1440

TTTAATTTAA AAACACAAAA AAAATTTTAG CTCTTCCCAC TTTTTTTTTC CTATTTATTT 1500

GAGGTCAGTG TTTGTTTTTG CACACCATTT TGTAAATGAA ACTTAAGAAT TGAATTGGAA 1560

AGACTTCTCA AAGAGAATTG TATGTAACGA TGTTGTWTTG ATTTTTAAGA AAGTAATTTA 1620

ATTTGTAAAA CTTCTGCTCG TTTACACTGC ACATTGAATA CAGGTAACTA ATTGGAAGGA 1680

GAGGGGAGGT CACTCTTTTG ATGGTGGCCC TGAACCTCAT TCTGGTTCCC TGCTGCGCTG 1740

CTTGGTGTGA CCCACGGAGG ATCCACTCCC AGGATGACGT GCTCCGTAGC TCTGCTGCTG 1800

ATACTGGGTC TGCGATGCAG CGGCGTGAGG CCTGGGCTGG TTGGAGAAGG TCACAACCCT 1860

TCTCTGTTGG TCTGCCTTCT GCTGAAAGAC TCGAGAACCA ACCAGGGAAG CTGTCCTGGA 1920

GGTCCCTGGT CGGAGAGGGA CATAGAATCT GTGACCTCTG ACAACTGTGA AGCCACCCTG 1980

GGCTACAGAA ACCACAGTCT TCCCAGCAAT TATTACAATT CTTGAATTCC TTGGGGATTT 2040

TTTACTGCCC TTTCAAAGCA CTTAAGTGTT AGATCTAACG TGTTCCAGTG TCTGTCTGAG 2100

GTGACTTAAA AAATCAGAAC AAAACTTCTA TTATCCAGAG TCATGGGAGA GTACACCCTT 2160

TCCAGGAATA ATGTTTTGGG AAACACTGAA ATGAAATCTT CCCAGTATTA TAAATTGTGT 2220

ATTTAAAAAA AAGAAACTTT TCTGAATGCC TACTGGCGGT GTATACCAGG CAGTGTGCCA 2280

GTTTAAAAAG ATGAAAAAGA ATAAAAACTT TTGAGGAAMA AAAAAAAAAA AAAAACTCGA 2340

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

AGTAGGGGAG ACTGAGACTG ACCGGTAGCC AGGCAGGCGG ACGACGCACG CCCGGACAGA 60

CTGAGCAGGC GCCGGAGAAC CACTCACAGG TTCCCCCCGC CTTTCCCTTT GAAANCTAGG 120

CTTTTGCCTT TCCCGTGGCG CCCGAGAGAG AATGCTGGAC TCTGCCGACT TCAGCGCAAC 180

TAANGATTTC TCAAGCTAGG GGACAAACGA TCAGCCCAAT CCTGAGAAGG GGGGAACCAA 240

GCACCCCGTC CCCATCCCCC TCCCCTCCCC CGACTAAACT CGGGCGCCAA ACCCAGCCCT 300

TCTCTAACCA CCCTACTTCC TCCTCTCCTT TCTAGCATGG TGGCTGTATG GACAGTCTGA 360

CAGAACAGAG ACTGACATCT CCCAATCTGC CGGCCCCCCA CCTGGAACAC TACAGTGTTC 420

TGCATTGCAC CATGACCCTG GATGTGCAAA CTGTAGTCGT TTTTGCCGTG ATTGTAGTCC 480

TCCTGCTTGT CAATGTCATA CTCATGTTTT TCCTGGGAAC GCGCTGAATG GAGTCCAGNC 540

ACCTGAGCTG TCGCGAACTC TCGCTTTGAT TTCATCCCGA GAGCCACCGA GAAGAAAAAA 600

A 601

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

CTCGTGCCGA ATTCGGCACG AGAGATGGTA CTTTTAAGAG GTAATTAGGT TGCTAAGATG 60

GATTAACATC TTTCTCTTGA CACTGAGACT GGGTTCTCCT GGGAATGGTT AGTTCCCAAG 120

AGAGTGAGTT GTTATAAAAC AATGCTGCCT CTTCTATTTT GCGCTTTTTG TTTGCACAAA 180

CTCGGTCCCC TTCTGTTTCT CTACGATGTT TTGATGCRGC ATGAGGCAGT CATGAGAACC 240

CACCAGATAC AGCTGCCTGA TCCTGAATTT CCCAGCCAAC AGAACCAAGT GCTAAATAAA 300

ACTCTTTTTA ATAAGTTAAA AAAAAAAAAA AAAAAAAAAA AANAAANANA AAAAAAAAA 359

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1141 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

GGCACGAGCT GCTGAGGCGT GAGAATGGCG TCCCGCGGCC GGCGTCCGGA GCATGGCGGA 60

CCCCCAGAGC TGTTTTATGA CGAGACAGAA GCCCGGAAAT ACGTTCGCAA CTCACGGATG 120

ATTGATATCC AGACCAGGAT GGCTGGGCGA GCATTGGAGC TTCTTTATCT GCCAGAGAAT 180

AAGCCCTGTT ACCTGCTGGA TATTGGCTGT GGCACTGGGC TGAGTGGAAG TTATCTGTCA 240

GATGAAGGGC ACTATTGGGT GGGCCTGGAT ATCAGCCCTG CCATGCTGGA TGAGGCTGTG 300

GACCGAGAGA TAGAGGGAGA CCTGCTGCTG GGGGATATGG GCCAGGGCAT CCCATTCAAG 360

CCAGGCACAT TTGATGGTTG CATCAGCATT TCTGCTGTGC AGTGGCTCTG TAATGCTAAC 420

AAGAAGTCTG AAAACCCTGC CAAGCGCCTG TACTGCTTTT TTGCTTCTCT TTTTTCTGTT 480

CTCGTCCGGG GATCCCGAGC TGTCCTGCAG CTGTACCCTG AGAACTCAGA GCAGTTGGAG 540

CTGATCACAA CCCAGGCCAC AAAGGCAGGC TTCTCCGGTG GCATGGTGGT AGACTACCCT 600

AACAGTGCCA AAGCAAAGAA ATTCTACCTC TGCTTGTTTT CTGGGCCTTC GACCTTTATA 660

CCAGAGGGGC TGAGTGAAAA TCAGGATGAA GTTGAACCCA GGGAGTCTGT GTTCACCAAT 720

GAGAGGTTCC CATTAAGGAT GTCGAGGCGG GGAATGGTGA GGAAGAGTCG GGCATGGGTG 780

CTGGAGAAGA AGGAGCGGCA CAGGCGCCAG GGCAGGGAAG TCAGACCTGA CACCCAGTAC 840

ACCGGCCGCA AGCGCAAGCC CCGCTTCTAA GTCACCACGC GGTTCTGGAA AGGCACTTGC 900

CTCTGCACTT TTCTATATTG TTCAGCTGAC AAAGTAGTAT TTTAGAAAAG TTCTAAAGTT 960

ATAAAAATGT TTTCTGCAGT AAAAAAAAAG TTCTCTGGGC CGGGCGTGGT GGCTCACACC 1020

TGTAATCCCA GCACCTTGGG AGGCTGAGGT GGGAGGATCA TTTGAGGCCA GGAGTTTGAG 1080

ACCTGCCTGG GCAACATAAT GAAACTTCCT TTCCAGGGAG AAAAAAAAAA AAAAAAAAAA 1140

A 1141

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

TCCTTCTCTG GGGCGGTCGC GTTGGCAGCG GATGCGGGAA GCCGGACTCT GGGCGTCATG 60

TACTACAAGT TTAGTGGCTT CACGCAGAAG TTGGCAGGAG CATGGGCTTC GGAGGCCTAT 120

AGCCCGCAGA TTNAAAGCCT GTGGTTTCCA CAGAAGCACC ACCTATCATA TTTGCCACAC 180

CAACTAAACT GACCTCCGAT TCCACAGTGT ATGATTATGC TGGGAAAAAC AAAGTTCCAG 240

AGCTACAAAA GTTTTTCCAG AAAGCTGATG GTGTGCCCGT CTACCTGAAA CGAGGCCTGC 300

CTGACCAAAT GCTTTACCGG ACCACCATGG CGCTGACTGT GGGAGGGACC ATCTACTGCC 360

TGATCGCCCT CTACATGGCT TCGCAGCCCA AAAACAAATG AGTTAGGCTG CAGAGGACTG 420

GTTTGTTTTT TGGCATAAAC CCTTTGAAGT TCCTTTTTCA TTGTTAAATT AAAATTTTTT 480

TTTTTACTTG GATGGCTTAA CATTTTTGCA AGAAAAATAG GAAGATATGA AGATGATGTT 540

TTGGTTTGTT TATGAAATGC ATATGGCTTG TCAGAGCTCA TTCGACAGTT AAAGCCATTG 600

TTTAAAGAAA CGGTGCTTTG CTCTGTGTTT GTGCTCCTGA TTTCCCTGGA GGTTCTGGAT 660

GAAGGCTGAA CACAGGCTTG TTAATGTCAG TCTGTGCTGA GGACCTCAGG GACTTGAGGT 720

TGCATTTTTG AGCATGGGGT GCAGGAGCCT TTCTGGATTT GGATGTGGCT ATGGAAAGAA 780

CACAGAAGCC AAGGTCATGT GCATGAAATG AGGAGTTTGA GTTAGTCACC TCGGGGATTT 840

TTTCCATTTT GCAGTAAAAT GTTAAATTAA TGTAGCCTGC CTCTATTTGT TGGGCAGGTA 900

ATTTCAAAGG GTTATTTGCC TCATCTCCTA TCTTTAGTGA AATCTTATGT GTAATTGTGT 960

GTATTTATTC CACCGTGGGA ACAGAGAATA CCTGTTTAGT GTTGCACTTT AGACTGGTGT 1020

CTGTTTTGTT AATGCAGCTG TGCCACAAAT TCTCCTTTAT CTTTTAAAAA TGTTATAGCT 1080

TTAAATTTTG ATTTATTTTG ACTGTGGAAT AAATACATGA ATGAAAAATT TTAAGTTTGA 1140

AGTTCTTTGA ATGACCTTTC AGAGTAATTT CAGAACACCA GCAGCATCTT AAACCTGAGT 1200

CTAATTTCTT TCTTGTTAAT TAGGCACCAG ATAATCTTTA TAAAATGGTC TTAAAAGCTA 1260

GTAATAGGAG CTTAATGGCA ATKGATGATT ACCACAKGGT TTTTTATAAA AACCTGCCTG 1320

CCCCTWAGTG AAAGGTACCT GTAACYCACA GTYCATTTAG ACACTAATTT CCTYTGCYGT 1380

CATGATTGGK AGACTTCACT TACCCTATAT TAATTTTGAA AAAAGGTGGA ATTTTATTAT 1440

ATATGAAGGA ATAGTTTGTA TCTTACCATA GCACAGAACA GTGACCTCTT GCTCAGGATA 1500

AGATGTGGTG ATTTGAAAAT ACTCATAGTA GCCTTGCAGT GATACCTCTC TCNCTCTCTC 1560

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1507 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

GGAACGCAGA GCGGAGCGTG GAGAGCGGAG CGAAGCTGGA TAACAGGGGA CCGATGATGT 60

GGCGACCATC AGTTCTGCTG CTTCTGTTGC TACTGAGGCA CGGGGCCCAG GGGAAGCCAT 120

CCCCAGACGC AGGCCCTCAT GGCCAGGGGA GGGTGCACCA GGCGGCCCCC CTGAGCGACG 180

CTCCCCATGA TGACGCCCAC GGGAACTTCC AGTACGACCA TGAGGCTTTC CTGGGACGGG 240

AAGTGGCCAA GGAATTCGAC CAACTCACCC CAGAGGAAAG CCAGGCCCGT CTGGGGCGGA 300

TCGTGGACCG CATGGACCGC GCGGGGGACG GCGACGGCTG GGTGTCGCTG GCCGAGCTTC 360

GCGCGTGGAT CGCGCACACG CAGCAGCGGC ACATACGGGA CTCGGTGAGC GCGGCCTGGG 420

ACACGTACGA CACGGACCGC GACGGGCGTG TGGGTTGGGA GGAGCTGCGC AACGCCACCT 480

ATGGCCACTA CGCGCCCGGT GAAGAATTTC ATGACGTGGA GGATGCAGAG ACCTACAAAA 540

AGATGCTGGC TCGGGACGAG CGGCGTTTCC GGGTGGCCGA CCAGGATGGG GACTCGATGG 600

CCACTCGAGA GGAGCTGACA GCCTTCCTGC ACCCCGAGGA GTTCCCTCAC ATGCGGGACA 660

TCGTGATTGC TGAAACCCTG GAGGACCTGG ACAGAAACAA AGATGGCTAT GTCCAGGTGG 720

AGGAGTACAT CGCGGATCTG TACTCAGCCG AGCCTGGGGA GGAGGAGCCG GCGTGGGTGC 780

AGACGGAGAG GCAGCAGTTC CGGGACTTCC GGGATCTGAA CAAGGATGGG CACCTGGATG 840

GGAGTGAGGT GGGCCACTGG GTGCTGCCCC CTGCCCAGGA CCAGCCCCTG GTGGAAGCCA 900

ACCACCTGCT GCACGARAGC GACACGGACA AGGAYGGGCG GCTGAGCAAA GCGSAAATCC 960

TGGGTAATTG GAACATGTTT GTGGGCAGTC AGGCCACCAA CTATGGYGAG GACCTGACCC 1020

GGCACCACGA TGAGCTGTGA GCMCCGNGCA CCTGCCACAG CCTCAGAGGC CCGCACAATG 1080

ACCGGAGGAG GGGCCGCTGT GGTCTGGCCC CCTCCCTGTC CAGGCCCCGC AGGAGGCAGA 1140

TGCAGTCCCA GGCATCCTCC TKCCCCTGGG CTCTCAGGGA CCCCCTGGGT CGGCTTCTGT 1200

CCCTGTCACA CCCCCAACCC CAGGGAGGGG CTGTCATAGT CCCAGAGGAT AAGCAATACC 1260

TATTTCTGAC TGAGTCTCCC AGCCCAGACC CAGGGACCCT NGGCCCCAAG CTCAGCTCTA 1320

AGAACCGCCC CAACCCCTCC AGCTCCAAAT CTGAGCCTCC ACCACATAGA CTGAAACTCC 1380

CCTGGCCCCA GCCCTCTCCT GCCTGGCCTG GCCTGGGACA CCTCCTCTCT GCCAGGAGGC 1440

AATAAAAGCC AGCGCCGGGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1500

AAAAAAN 1507

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 450 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

GAATTCGGCA CGAGCAGTGT CCAACACTGT AGCTGGTGCC TGCCAGGTTC CCAGTGGCTG 60

GGGTCACCAG GTCTGAAGAG AGATGTGCTG GCTGCGGGCA TGGGSCCAGA TCYTCCTGCC 120

AGTTTTCYTC TCCYTCTTTC TCATCCAATT GCTTATCAGC TTCTCAGAGA ATGGTTTTAT 180

CCACAGCCCC AGGAACAATC AGAAACCAAG AGATGGGAAT RAAGAGGAAT GTGCTGTAAA 240

GAAGAGTTGT CAATTGTGCA CAGAAGATAA GAAATATATG ATGAATAGAT AATTGAAAAG 300

AGATCCTCCA GAAAGAGCAG AAGGAAGTTT CTTCAATGGC TTCCTTCAGG ATTTTAATCA 360

TCCTTACAGC CTCTTTGAGA ATGATTGAAC TTCCAAATTC CCTGAAGTTA AAATTTTAAA 420

TTCTATTAAA CATTTTTTCG AGTAAAAAAA 450

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1147 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

GGCACGAGAC CCATTGAGCA GAAGGAGGCC AGGTGGGAAA GCTCCTGGGA AGAGCAGCCA 60

GACTGGACAC TGGGCTGCTT GAGTCCTGAG TCACAATTCA GAATTCCTGG GCTCCCTGGG 120

TGCATTCTAT CATTCCAGTT GAAAGTTTGC TTCCTTCCAG TCATGTGGCT CTTCATTCTA 180

CTCTCCTTGG CTCTCATTTC AGATGCCATG GTCATGGATG AAAAGGTCAA GAGAAGCTTT 240

GTGCTGGACA CGGCTTCTGC CATCTGCAAC TACAATGCCC ACTACAAGAA TCACCCCAAA 300

TACTGGTGCC GAGGCTATTT CCGTGACTAC TGCAACATCA TCGCCTTCTC CCCTAACAGC 360

ACCAATCATG TGGCCCTGAA GGACACAGGG AACCAGCTCA TTGTCACTAT GTCCTGCCTG 420

AACAAAGAAG ACACGGGCTG GTACTGGTGT GGCATCCAGC GGGACTTTGC CAGGGATGAC 480

ATGGATTTTA CAGAGCTGAT TGTAACTGAC GACAAAGGAA CCTGGCCAAT GACTTTGGTC 540

TGGGAAAGAC TATCAGGCAC AAAACCAGAA GCTGCAAGGC TCCCAAAGTT GTCCGCAAGG 600

CTGACCGCTC CAGGACGTCC ATTCTCATCA TTTGCATACT GATCACGGGT TTGGGAATCA 660

TCTCTGTAAT CAGTCATTTG ACCAAAAGGA GGAGAAGTCA AAGGAATAGA AGGGTAGGCA 720

ACACTTTGAA GCCCTTCTCG CGTGTCCTGA CTCCAAAGGA AATGGCTCCT ACTGAACAGA 780

TGTGACTGAA GATTTTTTTA ATTTAGTTCA TAAAGTGATG CTACAACAGA ATAATCACCA 840

TGACAACTGG CCCCACACCT CAGAGACTGA TTCTGATCTC CCAGGAATTC TGAAGGTCCC 900

TCTATCCTTG ACAACAATCA TTTGCAGCCA GGTAGCAACG GCAGTAGTCA GAGGAGCTAT 960

GATAGACCAC ACCCAAGCAA GGCTGCCCTC AAATAACATC TCAAGATCTT AGTTCTTATG 1020

CATTCCATCA GTCAGAAGTG AAGAAGAGGT GGAGAATCTG GATTGGGGAC CAGGAAATCA 1080

CTTGTATTTT GTTAGCCAAT AAATTCCTAG CCAGTGTTGA ATGAAAAAAA AAAAAAAAAA 1140

AAAAAAA 1147

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 777 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

GGCAGAGGCT CCTCAGAAGG GCGTGGGCTC TCCAGTCTTC CACAGTCCCC ACCATGCCCT 60

GTTGCCTTAC CGCTGACGTA GCTCACCCAT CTTTTACTTG CCTGGCTAAG ATGCATGGCA 120

TYWCATTTCC TCCTTGTTGC ACTGCAGTCA GTCCCTCACT GCCCCCATCT CCTGGAAGAG 180

GAGCATAAGC TTTGCAAGGT CAGCCACTTC TCTGGGGTCA CACTAGTTAC ATCAAGACAG 240

GACTCCAGCT CATATGTGCC AGTGCAGACA CTCTTCATCC ACCTGGGGCC CTGGGCTTGG 300

GACCTGGYTC CTTGCACAGC AGARGACCCG GAGGCTGAGA GGAGCTTGCG GTTGTGTCAT 360

AGTCACCTGG CCAGARGGAA CGTGAGCCCC TCCCAAGCTG CAGARGGARG GARCARGCGT 420

GGCTGTCAGC ACCGAGGTAG CAGAGAATTA ACATTCTTGT CAGCAGAGAA TGAAGCAGGA 480

ATATAATTAA AACTTTGCCC TTGGAATAGC TGATTCATTT GAATTTTATT CCACACGTTT 540

GAAAGAGGAA AGAAAATGTG AAGACTTGCA GCCTGGTTCT CGCCTGGCCT GGGCTGGCCC 600

AGCTGTCAGG CCCGGTTCCT TTCTGAGCAT TCAGTCCACT GATGTTGACT GAGGGCCAGG 660

AGAGACCCTC AGCAGGGTAT TACCATATCA GCCTCCTATC GCTGCTGGGA GAAATTACCA 720

TGAATTCAGT GGCTTAAAAC AACACACGAG CCTCTCTGAG CCTACCCTGG CTCAGGA 777

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1191 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

AAGANTGATT TTCCTTACTC TCCAAAGCGT CAGCATTTTG AAGTTTCTTT TATGAAAGTG 60

GGGGCAAGAA TCAGGGTGAA AATGAGTGTA AACAAAGCCC ATCCTGTGGT CAGCACCCAC 120

TGGAGGTGGC CAGCAGAGTG GCCTCAGATG TTCCTGCACC TGGCCCAGGA GCCCAGGACA 180

GAGGTCAAAT CTAGGCCCCT TGGTCTGGCT GGATTCATCA GGCAAGATTC GAAAACAAGA 240

AAACCTCTAG AACAAGAAAC AATCATGTCT GCAGCAGATA CGGCACTGTG GCCCTATGGC 300

CATGGCAATC GTGAGCACCA AGAGAATGAG TTACAGAAAT ATCTCCAATA CAAAGACATG 360

CATCTCCTGG ACAGTGGACA GTCGCTGGGA CACACACACA CACTTCAAGG CTCACACAAC 420

CTAACAGCCT TAAATATCTG AAGAAACAGA ATCACGACAT TAAGTCAGCA GAGGGAGAGG 480

TAGGCTGAAG CAGCAGGAGG CCAATTTTAT ATCCCACAGA TTTTTTTAAA AATGACTCCC 540

CAGCAAGGGG TGGGGAGAAA GCCACTGATT TAGGAGAGTT CTTGGCTCAG CCAACCACTG 600

CGGTTATCTA CACGTTTTAC AAAGGCACRG AAGTAGAGAG GGGCTGCACT CACGACCCTC 660

CCCAGGGCCC GCACAGCCAG ACACGGTGGG TTCTTCCTTT TTCCCTTCTG GCCTTGGTGG 720

AATTCCTACC ACGGTGGCCT CTGCCTTTGG GACAATGCCT TCATGCTCAT CCCCGGGTCA 780

AGGATGGAGT CTGTTACCAT TTTCCAGGGG AAATTCCAAG GACCAGCCCC GCCTCATTAC 840

GTTCACCCCA CAGGAAGGTG ATCTGGAAAG CCTGTAAACA CGTACTCTGG GTGGCTGAGT 900

GGTGTCACCA AGCTGCTTTT GTGCAGGGCT GAAGCACAGA CAAGAGGGCA GGCAGCTGCC 960

GGAGGCCTGA AGTGGGGAGA GATCCCCGCA GGCCTGCAGG AGCCAGGGAG AACCTCCAAC 1020

TGGATCTAAA CTGTGGGACA GCCCAGGCGT GCCCCTCTTC ACATGGCTCC CAGGCTCCCT 1080

CAAAGCCCTT CCCAGGCCCT GCAGGAAGAG AGGGAGGGTG AGGAGAGGCA GGGAGGGCAG 1140

AGGTCGCCTG AAAGCCTGGG CTCCGAACTC CCTCAGCAGA GCTTTAAAGT G 1191

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CCCCGCCCCC CGCCCACGAA GGAAGTGGCT GCTGCTCCGG CGCGGACCCA GAGCCGGTTC 60

GGCGCGTCGA CTGCCCAGAG TCCGCGGCCG GGCGCGGGAG GAGCCAAGCC GCCATGGCCT 120

ACCACAGCTT CCTGGTGGAG CCCATCAGCT GCCACGCCTG GAACAAGGAC CGCACCCAGA 180

TTGCCATCTG CCCCAACAAC CATGAGGTGC ATATCTATGA AAAGAGCGGT GCCAAATGGA 240

CCAAGGTGCA CGAGCTCAAG GAGCACAACG GGCAGGTGAC AGGCATCGAC TGGGCCCCCG 300

AGAGTAACCG TATTGTGACC TGCGGCACAG ACCGCAACGC CTACGTGTGG ACGCTGAAGG 360

GCCGCACATG GAAGCCCACG CTGGTCATCC TGCGGATCAA CCGGGCTGCC CGCTGCGTGC 420

GCTGGGCCCC CAACGAGAAC AAGTTTGCTG TGGGCAGCGG CTCTCGTGTG ATCTCCATCT 480

GTTATTTCGA GCAGGAGAAT GACTGGTGGG TTTGCAAGCA CATCAAGAAG CCCATCCGCT 540

CCACCGTCCT CAGCCTGGAC TGGCACCCCA ACAATGTGCT GCTGGCTGCC GGCTCCTGTG 600

ACTTCAAGTG TCGGATCTTT TCAGCCTACA TCAAGGAGGT GGAGGAACGG CCGGCACCCA 660

CCCCGTGGGG CTCCAAGATG CCCTTTGGGG AACTGATGTT CGAATCCAGC AGTAGCTGCG 720

GCTGGGTACA TGGCGTCTGT TTCTCAGCCA GCGGGAGCCG CGTGGCCTGG GTAAGCCACG 780

ACAGCACCGT CTGCCTGGCT GATGCCGACA AGAAGATGGC CGTCGCGACT CTGGCCTCTG 840

AAACACTACC ACTGCTGGCG CTGACCTTCA TCACAGACAA CAGCCTGGTG GCAGCGGGCC 900

ACGACTGCTT CCCGGTGCTG TTCACCTATG ACGCCGCCGC GGGGATGCTG AGCTTCGGCG 960

GGCGGCTGGA CGTTCCTAAG CAGAGCTCGC AGCGTGGCTT GACGGCCCGC GAGCGCTTCC 1020

AGAACCTGGA CAAGAAGGCG AGCTCCGAGG GTGGCACGGC TGCGGGCGCG GGCCTAGACT 1080

CGCTGCACAA GAACAGCGTC AGCCAGATCT CGGTGCTCAG CGGCGGCAAG GCCAAGTGCT 1140

CGCAGTTCTG CACCACTGGC ATGGATGGCG GCATGAGTAT CTGGGATGTG AAGAGCTTGG 1200

AGTCAGCCTT GAAGGACCTC AAGATCAAAT GACCTGTGAG GAATATGTTG CCTTCATCCT 1260

AGCTGCTGGG GAAGCGGGGA GAGGGGTCAG GGAGGCTAAT GGTTGCTTTG CTGAATGTTT 1320

CTGGGGTACC AATACGAGTT CCCATAGGGG CTGCTCCCTC AAAAAGGGAG GGGACAGATG 1380

GGGAGCTTTT CTTACCTATT CAAGGAATAC GTGCCTTTTT CTTAAATGCT TTCATTTATT 1440

GAAAAAAAAA AAAAATGCCC CCAAAGCACT ATGCTGGTCA TGAACTGCTT CAAAATGTGG 1500

AGGTAATAAA ATGCAACTGT GTAAAAAAAA AAAAAAAAAA AAATGACCCT CGCGATCTAG 1560

AACTAGNCGG ACGCNTGGGT 1580

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1117 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

GGCACGAGGC GCGATGCAGC ACAGGCTAGA GGCTGCGCAA SGCGGGGGCC CGCCCCTGGG 60

ACCCTCCGGG CCGGGCGGTT TGGCCCCTTA GCGCCCGGGC GTCGGGGCGG TAAAAGGCCG 120

GCAGAAGGGA GGCACTTGAG AAATGTCTTT CCTCCAGGAC CCAAGTTTCT TCACCATGGG 180

GATGTGGTCC ATTGGTGCAG GAGCCCTGGG GGCTGCTGCC TTGGCATTGC TGCTTGCCAA 240

CACAGACGTG TTTCTGTCCA AGCCCCAGAA AGCGGCCCTG GAGTACCTGG AGGATATAGA 300

CCTGAAAACA CTGGAGAAGG AACCAAGGAC TTTCAAAGCA AAGGAGCTAT GGGAAAAAAA 360

TGGAGCTGTG ATTATGGCCG TGCGGAGGCC AGGCTGTTTC CTCTGTCGAG AGGAAGCTGC 420

GGATCTGTCC TCCCTGAAAA GCATGTTGGA CCAGCTGGGC GTCCCCCTCT ATGCAGTGGT 480

AAAGGAGCAC ATCAGGACTG AAGTGAAGGA TTTCCAGCCT TATTTCAAAG GAGAAATCTT 540

CCTGGATGAA AAGAAAAAGT TCTATGGTCC ACAAAGGCGG AAGATGATGT TTATGGGATT 600

TATCCGTCTG GGAGTGTGGT ACAACTTCTT CCGAGCCTGG AACGGAGGCT TCTCTGGAAA 660

CCTGGAAGGA GAAGGCTTCA TCCTTGGGGG AGTTTTCGTG GTGGGATCAG GAAAGCAGGG 720

CATTCTTCTT GAGCACCGAG AAAAAGAATT TGGAGACAAA GTAAACCTAC TTTCTGTTCT 780

GGAAGCTGCT AAGATGATCA AACCACAGAC TTTGGCCTCA GAGAAAAAAT GATTGTGTGA 840

AACTGCCCAG CTCAGGGATA ACCAGGGACA TTCACCTGTG TTCATGGGAT GTATTGTTTC 900

CACTCGTGTC CCTAAGGAGT GAGAAACCCA TTTATACTCT ACTCTCAGTA TGGATTATTA 960

ATGTATTTTA ATATTCTGTT TAGGCCCACT AAGGCAAAAT AGCCCCAAAA CAAGACTGAC 1020

AAAAATCTGA AAAACTAATG AGGATTATTA AGCTAAAACC TGGGAAATAG GAGGCTTWAA 1080

ATGACTGCCM GCTGGTGCRT GCTCACACTT GGCCCAC 1117

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 361 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CCCACGCGTG CKGGCGCCTG GCAGCCACCG CCTGGGAGGT TACTGTAAGG CCCGCAGCTC 60

CCGCCAGCTC CCGCGGACTS CTGCCGCCTC CTTACCATGA AGCCAGTAAG TCGTCGCACG 120

CTGGACTGGA TTTATTCAGT GTTGCTGCTT GCCATCGTTT TAATCTCCTG GGGCTGCATC 180

ATCTATGCTT CGATGGTGTC TGCAAGACGA CAGCTAAGGA AGAAATACCC AGACAAAATC 240

TTTGGGACGA ATGAAAATTT GTAACTCTTC TGGATTTAAT TATCTGAAAA TACAGTTCTT 300

TCCCTCATGC TTATGTAGAT ATAAAAATAA AATTCATAAT GCAAAAAAAA AAAAAAAAAA 360

G 361

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1668 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

GGCACGAGGT CTGCCAAGCT ATAGACCATG GCTGTGAACA CATTTGTGTG AACAGTGACG 60

ACTCATACAC GTGCGAGTGC TTGGAGGGAT TCCGGCTCGC TGAGGATGGG AAACGCTGCC 120

GAAGAAGGAT GTCTGCAAAT CAACCCACCA TGGCTGCGAA CACATTTGTG TTAATAATGG 180

GAATTCCTAC ATCTGCAAAT GCTCAKAGGG ATTTGTTCTA GCTGAGGACG GAAGACGGTG 240

CAAGAAATGC ACTGAAGGCC CAATTGACCT GGTCTTTGTG ATCGATGGAT CCAAGAGTCT 300

TGGAGAAGAG AATTTTGAGG TCGTGAAGCA GTTTGTCACT GGAATTATAG ATTCCTTGAC 360

AATTTCCCCC AAAGCCGCTC GAGTGGGGCT GCTCCAGTAT TCCACACAGG TCCACACAGA 420

GTTCACTCTG AGAAACTTCA ACTCAGCCAA AGACATGAAA AAAGCCGTGG CCCACATGAA 480

ATACATGGGA AAGGGCTCTA TGACTGGGCT GGCCCTGAAA CACATGTTTG AGAGAAGTTT 540

TACCCAAGGA GAAGGGGCCA GGCCCTTTCC ACAAGGGTGC CCAGAGCAGC CATTGTGTTC 600

ACCGACGGAC GGGCTCAGGA TGACGTCTCC GAGTGGGCCA GTAAAGCCAA GGCCAATGGT 660

ATCACTATGT ATGCTGTTGG GGTAGGAAAA GCCATTGAGG AGGAACTACA AGAGATTGCC 720

TCTGAGCCCA CAAACAAGCA TCTCTTCTAT GCCGAAGACT TCAGCACAAT GGATGAGATA 780

AGTGAAAAAC TCAAGAAAGG CATCTGTGAA GCTCTAGAAG ACTCCGATGG AAGACAGGAC 840

TCTCCAGCAG GGGAACTGCC AAAAACGGTC CAACAGCCAA CAGTGCAACA CAGATATCTG 900

TTTGAAGAAG ACAATCTTTT ACGGTCTACA CAAAAGCTTT CCCATTCAAC AAAACCTTCA 960

GGAAGCCCTT TGGAAGAAAA ACACGATCAA TGCAAATGTG AAAACCTTAT AATGTTCCAG 1020

AACCTTGCAA ACGAAGAAGT AAGAAAATTA ACACAGCGCT TAGAAGAAAT GACACAGAGA 1080

ATGGAAGCCC TGGAAAATCG CCTGAGATAC AGATGAAGAT TAGAAATCGC GACACATTTG 1140

TAGTCATTGT ATCACGGATT ACAATGAACG CAGTGCAGAG CCCCAAAGCT CAGGCTATTG 1200

TTAAATCAAT AATGTTGTGA AGTAAAACAA TCAGTACTGA GAAACCTGGT TTGCCACAGA 1260

ACAAAGACAA GAAGTATACA CTAACTTGTA TAAATTTATC TAGGAAAAAA ATCCTTCAGA 1320

ATTCTAAGAT GAATTTACCA GGTGAGAATG AATAAGCTAT GCAAGGTATT TTGTAATATA 1380

CTGTGGACAC AACTTGCTTC TGCCTCATCC TGCCTTAGTG TGCAATCTCA TTTGACTATA 1440

CGATAAAGTT TGCACAGTCT TACTTCTGTA GAACACTGGC CATAGGAAAT GCTGTTTTTT 1500

TGTAYTGGAC TTTACCTTGA TATATGTATA TGGATGTATG CATAAAATCA TAGGACATAT 1560

GTACTTGTGG AACAAGTTGG ATTTTTTATA CAATATTAAA ATTCACCACT TCAGAGRAAA 1620

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAANAAAA 1668

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1353 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

GGGTCGACCC ACGCGTCCGC CCACGCGTCC GGATGGCTGC GCTGTTGCTG AGACACGTTG 60

GTCGTCATTG CCTCCGAGCC CACTTTAGCC CTCAGCTCTG TATCAGAAAT GCTGTTCCTT 120

TGGGAACCAC GGCCAAAGAA GAGATGGAGC GGTTCTGGAA TAAGAATATA GGTTCAAACC 180

GTCCTCTGTC TCCCCACATT ACTATCTACA GTTGGTCTCT TCCCATGGCG ATGTCCATCT 240

GCCACCGTGG CACTGGTATT GCTTTGAGTG CAGGGGTCTC TCTTTTTGGC ATGTCGGCCC 300

TGTTACTCCC TGGGAACTTT GAGTCTTATT TGGAACTTGT GAAGTCCCTG TGTCTGGGGC 360

CAGCACTGAT CCACACAGCT AAGTTTGCAC TTGTCTTCCC TCTCATGTAT CATACCTGGA 420

ATGGGATCCG ACACTTGATG TGGGACCTAG GAAAAGGCCT GAAGATTCCC CAGCTATACC 480

AGTCTGGAGT GGTTGTCCTG GTTCTTACTG TGTTGTCCTC TATGGGGCTG GCAGCCATGT 540

GAAGAAAGGA GGCTCCCAGC ATCATCTTCC TACACATTAT TACATTCACC CATCTTTCTG 600

TTTGTCATTC TTATCTCCAG CCTGGGAAAA GTTCTCCTTA TTTGTTTAGA TCCTTTTGTA 660

TTTTCAGATC TCCTTGGAGC AGTAGAGTAC CTGGTAGACC ATAATAGTGG AAAAGGGTCT 720

AGTTTTCCCC TTGTTTCTAA AGATGAGGTG GCTGCAAAAA CTCCCCTTTT TTGCCCACAG 780

CTTGCCTACT CTCGGCCTAG AAGCAGTTAT TCTCTCTCCA TATTGGGCTT TGATTTGTGC 840

TGAGGGTCAG CTTTTGGCTC CTTCTTCCTG AGACAGTGGA AACAATGCCA GCTCTGTGGC 900

TTCTGCCCTG GGGATGGGCC GGGTTGGGGG GTGGGTTGGT GAGGCTTTGG GTGCCACTGC 960

CTGTGGGTTG CTGGCTTAAA GGACAATTCT CTTCATTGGT GAGAGCCCAG GCCATTAACA 1020

CCTACACAGT GTTATTGAAA GAAGAGAGGT GGGGGTGGAG GGGAATTAGT CTGTCCCAGC 1080

TAGAGGGAGA TAAAGAGGGC TAGTTAGTTC TTGGAGCAGC TGCTTTTGAG GAGAAAATAT 1140

ATAGCTTTGG ACACGAGGAA GATCTAGAAA ATTATCATTG AACATATTAA TGGTTATTTC 1200

TTTTTCTTGG ATTTCCAGAA AAGCCTCTTA ATTTTATGCT TTCTCATCGA AGTAATGTAC 1260

CCTTTTTTTC TGAAACTGAA TTAAATACTC ATTTTATCTT TGAAAAAAAA AAAAAAAACC 1320

TNGGGGGGGG CCCCGGACCC NAATTGGCCC TAT 1353

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1011 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CGGAAGAAAG CAGCCATCCA GACATTTCAG AACACGTACC AGGTGTTAGC TGTGACCTTC 60

AATGACACAA GTGATCAGAT TATTTCTGGT GGAATAGACA ATGATATCAA GGTCTGGGAC 120

TGCGCCAGAA CAAGCTAACC TACACCATGA GAGGCCATGC AGATTCAGTG ACTGGCCTGA 180

GTTTAAGTTC TGAAGGCTCT TATCTTTTGT CCAATGCAAT GGACAATACA GTTCGTGTCT 240

GGGATGTCCG GCCATTTGCC CCCAAAGAGA GATGTGTAAA GATATTTCAA GGAAATGTGC 300

ACAACTTTGA AAAGAACCTT CTGAGATGTT CTTGGTCACC TGATGGAAGC AAAATAGCAG 360

CTGGCTCAGC CGACAGGTTT GTTTATGTGT GGGATACCAC AAGCAGGAGA ATATTGTATA 420

AGCTGCCCGG CCATGCTGGC TCCATCAATG AAGTGGCTTT CCACCCTGAT GAGCCCATCA 480

TTATCTCAGC ATCGAGTGAC AAGAGACTGT ATATGGGAGA GATTCAGTGA AGATATGGAC 540

TGGAAGACTC CAAGGCCGCT TGTCTTTGAG ACCTCAGACT GCATAAGTGA TGCCAAATGT 600

TGGATGTCCA GGYTAGCACC CTCCCTTCAG ATGACCATTG CTAGCAAGAA ACAGGAGGCG 660

GTGGCCATAT TCCAAAAACC ACTTCTGTCC CATTTCACCA GGATGACTAA GGCAAGCTCC 720

CTGTGGCCTC TAAAAACCAC CTGCCAGATT TCAGGGACTG TTTTTTTTTT TCTTTTTCTT 780

TTTTCCTGTT TTCTAATGCA GGCCCAATGT GACAAATTTG TTGGTTGGGA TTTTTTTTTT 840

TTTTTGTAAC TGGCTTGTAT GATATTTTCT TTCTGTATTT CTCTATATCA TTTTGTATTA 900

AAAGCCAAAT AGATGCCTTT TTACAAGARM AAAAAAAAAA AAAAAAAAAA NNAAAAAAAA 960

CTGGGAGGGG GGGCCCGGTA CCCAAATCGC CGGATATGAT CGTAAACAAT C 1011

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1193 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

GGCCGGGCGG TGCGCACTGC GGGCGCATCC CTGCCCCGGC GCCGTCCGTG CCCGCGGGAC 60

CTGACAGCCG GGTCAGAGGG CGAACTGTGC TCAGGCCCGG GCTGGACGCA GAGCCAGAGC 120

TGTCCCCAGA GGAGCAGAGG GTCCTGGAAA GGAAGCTGAA AAAGGAACGG AAGAAAGAGG 180

AGAGGCAGCG TCTGCGGGAG GCAGGCCTTG TGGCCCAGCA CCCGCCTGCC AGGCGCTCGG 240

GGGCCGAACT GGCCTGGGAC TACCTCTGCA GATGGGCCCA AAAGCACAAG AACTGGAGGT 300

TTCAGAAGAC GAGGCAGACG TGGCTCCTGC TGCACATGTA TGACAGTGAC AAGGTTCCCG 360

ATGAGCACTT CTCCACCCTG CTGGCCTACC TGGAGGGGCT GCAGGGCCGG GCCCGAGAGC 420

TGACGGTGCA GAAGGCGGAA GCCTGATGCG GGAGCTGGAT GAGGAGGGCT CTGATCCCCC 480

CCTGCCGGGG AGGGCCCAGC GCATCCGACA GNTGCTGCAG CTGCTCTCCT AGTGGGTTCA 540

GCGCGGGGCG GGGCCGCTGC CCAGTGCAGG GCTGCCTCAG ACCACACAGG GTGCAGCTCC 600

TCCGGCGGTG GGGGCCGGGT TCACCAGCAG GGCAGCGGCT GAGCAAGGGC TTTCAGCTCC 660

TCCGGTGGTG GGGGCCGGGA TCACCAGCAC CAGAGCCTCG CAAGGGCCCC TTCCCTCCTC 720

CAGACCCTCC TTGGCCGGTG ACGCTGTGAC AGTGATGGCA GGTTCAGTGC CTTCAGCGCA 780

GAGCGTGGAT GCTCTGGAAT CACCCGGACC CCTGGCCTTG GAGGGACCCT CCAGCCCCAG 840

GAATCTGCTT TGGAGGGAAA TGTCTATTTT TCTACCGGGA ATATTTTAGA GATTGGGGCA 900

TGCTGGCTCC TCCCGCCAGC TGCAAACCTG CACCTTCCGC CTGATTCCCG ATCCCCCTGC 960

GTGGGCCGCA TTCCTGGTCC CCTGCCTGCG TCCATCGAGG GGCCTGGCTG TGGCCTGTTT 1020

TCCTTTGACC CCACACAGCG TCATTGCGGG TCATGGGGAG CCCCTGGTGG GAGCTTGTGG 1080

AGTCGGATCA CGTACCTGTG CAGAAACCGC CTCTGTGGCT GCATTTGAAA TAAAACCCGA 1140

CCCAGCAGCA AAAAAAAAAA AAAAAANCNC NAGGGGGGGC CCGGNACCCA ATT 1193

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

GAATTCGGCA CGAGTTGGCA CATGATGCAA AATGCATTTC TCAGAGTAGA TTGCAGTCAA 60

AAATGTTGGA AACTACTAAG CATGTGCARA TAGCATGCAT GCTGCTGCTG ACCTGCCAGA 120

TATTTCTCCC TTCCTCCCTT TCTCCCTCAT TTATTCATTC ATTAACTGAT TCATTCATCC 180

CATTAAAAAA ATTATATGTA TGTTTTGTGC AAAGCACCCT ACTCAAGGCT GCGGGGTACA 240

AAAGTATATC AGAAGCCTTG GGCTTTGACM WACTTCTCTG TAGTAGTGCT AGATTTGTGT 300

GGATCTGCCA CACTTACTCC AGGCCTCTTG TGACCTGTGC TTTGCATTAA TCTCTTAGGC 360

TAAGCCACAT ACCTTTTCAT TATACAATCT TTGCTGATGC TAAGGACAGA TTCCAAAGTG 420

CCCTCCTTAT AATTTTTGTA TTTAATGCAA AGTGTAATCA AGAATAGGCC ATTGTTAGGT 480

CAATTGCTTT TCTGTATTTA TCTTTTCAAA CAATAAATAA TCAGTGGGAT GAAAAAGGGC 540

CGGAAAAAAA AAAAAAAAAA 560

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1657 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

```
CGGACNGAGC CGCCGCCGGG CACTTCCTGT GGAGGCCGCA GCGGGTGCGG GCGCCGACGG     60
GCGAGAGCCA GCGAGCGAGC GAGCGAGCCG AGCCGAGCCT CCCGCCGTCG CCATGGGCCA    120
GAACGACCTG ATGGGCACGG CCGAGGACTT CGCCGACCAG TTCCTCCGTG TCACAAAGCA    180
GTACCTGCCC CACGTGGCGC GCCTCTGTCT GATCAGCACC TTCCTGGAGG ACGGCATCCG    240
TATGTGGTTC CAGTGGAGCG AGCAGCGCGA CTACATCGAC ACCACCTGGA ACTGCGGCTA    300
CCTGCTGGCC TCGTCCTTCG TCTTCCTCAA CTTGCTGGGA CANTGACTGG CTGCGTCCTG    360
GTGTTGAGCA GGAACTTCGT GCAGTACGCC TGCTTCGGGC TCTTTGGAAT CATAGCTCTG    420
CAGACGATTG CCTACAGCAT TTTATGGGAC TTGAAGTTTT TGATGAGGAA CCTGGCCCTG    480
GGAGGAGGCC TGTTGCTGCT CCTAGCAGAA TCCCGTTCTG AAGGGAAGAG CATGTTTGCG    540
GGCGTCCCCA CCATGCGTGA GAGCTCCCCC AAACAGTACA TGCAGCTCGG AGGCAGGGTC    600
TTGCTGGTTC TGATGTTCAT GACCCTCCTT CACTTTGACG CCAGCTTCTT TTCTATTGTC    660
CAGAACATCG TGGGGCACAG CTCTGATGAT TTTAGTGGCC ATTGGTTTTA AAACCAAGCT    720
GGCTGCTTTG ACTCTTGTTG TGTGGCTCTT TGCCATCAAC GTATATTTCA ACGCCTTCTG    780
GACCATTCCA GTCTACAAGC CCATGCATGA CTTCCTGAAA TACGACTTCT TCCAGACCAT    840
GTCGGTGATT GGGGGCTTGC TCCTGGTGGT GGCCCTGGGC CCTGGGGGTG TCTCCATGGA    900
TGAGAAGAAG AAGGAGTGGT AACAGTCACA GATCCCTACC TGCCTGGCTA AGACCCGTGG    960
CCGTCAAGGA CTGGTTCGGG GTGGATTCAA CAAAACTGCC AGCTTTTATG TATCCTCTTC   1020
CCTTCCCCTC CCTTGGTAAA GGCACAGATG TTTTGAGAAC TTTATTTGCA GAGACACCTG   1080
AGAATCAATG GCTTCAGGAC ATGGGTTCTC TTCTCCTGTG ATCATTCAAG TGCTCACTGC   1140
ATGAAGACTG GCTTGTCTCA GTGTTTCAAC CTCACCAGGG CTGTCTCTTG GTCCACACCT   1200
CGCTCCCTGT TAGTGCCGTA TGACAGCCCC CATCAAATGA CCTTGGCCAA GTCACGGTTT   1260
CTCTGTGGTC AAGGTTGGTT GGCTGATTGG TGGAAAGTAG GGTGGACCAA AGGAGGCCAC   1320
GTGAGCAGTC AGCACCAGTT CTGCACCAGC AGCGCCTCCG TCCTAGTGGG TGTTCCTGTT   1380
TCTCCTGGCC CTGGGTGGGC TAGGGCCTGA TTCGGGAAGA TGCCTTTGCA GGGAGGGGAG   1440
GATAAGTGGG ATCTACCAAT TGATTCTGGC AAAACAATTT CTAAGATTTT TTTGCTTTAT   1500
GTGGGAAACA GATCTAAATC TCATTTTATG CTGTATTTTA TATCTTAGTT GTGTTTGAAA   1560
ACGTTTTGAT TTTTGGAAAC ACATCAAAAT AAATAATGGC GTTTGTTGTA AAAAAAAAAA   1620
```

AAAAAAACTC GRGGGGGGGC CCGGTACCCA AATCGCC 1657

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 711 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

GGCACGAGCG AAGACCCTGT TCGGACCCTG CCCCGATTCC AGACTCAGGT AGATCGTCGG 60

CATACCCTCT ACCGTGGACA CCAGGCAGCC CTGGGGCTGA TGGAGAGAGA TCAGGTATCC 120

CCCAGGGAGT AGGGGCTACC TTGAGGGGAT GATAGACCTC CCCCACTCCC AGTGKKACTC 180

TGGAAATATG AAGGAACTAG GGAGTGGAAG AGATTTCAGA GCTGGGGAGA GGAGTTCCTC 240

CCTTCAAAGC CAGCAACTGC CTTTGGGGAA TGTCGGGGGG TCTCTCCTTT CTCCTGCTTG 300

TTTRAGGTGG TACACAGTCC CCCCTTCAMC TGGSGGGAAG CTGTNCCGGA CARACTCATC 360

TCAGCTTTCC CTTGGGGCAG GATCGGGGGC AGCAGCTCCA GCAGAAACAG CAGGATCTGG 420

AGCAGGAAGG CCTCGAGGCC ACACAGGGGC TGCTGGCCGG CGAGTGGGCC CCACCCCTCT 480

GGRAGCTGGG CAGCCTCTTC CAGGCCTTCG TGAAGAGGGA GAGCCAGGCT TATGCGTAAG 540

CTTCATAGCT TCTGCTGGCC TGGGGTGGAC CCAGGACCCC TGGGGCCTGG GTGCCCTGAG 600

TGGTGGTAAA GTGGAGCAAT CCCTTCACGC TCCTTGGCCA TGTTCTGAGC GGCCAGCTTG 660

GCCTTTGCCT TAATAAATGT GCTTTATTTT CAAAAAAAAA AAAAAAAAAC T 711

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 935 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

GGCACAGGGT GAAAGCCAGC TAAACCCCAA GTGGAGAAGT GAAAGACATG GTTGTTCCCA 60

TAAGTTTATT GCTCACATTA TGAAAGAAGC CATAGTCATG AGTGAACCAC TCCCTAGGTT 120

GATAAGGAAA CCAACACGGA AGATCTCTTT CTGGAAGAAG CAGCCAGCCT CGTGAAGGAG 180

CGGCCCAGCC GCCGGGCCCG AGGGTCGCCT TTTGTTCGGA GTGGCACGAT TGTCCGTTCC 240

CAGACATTCT CGCCTGGAGC ACGAAGCCAG TATGTTTGCA GACTTTATCG TAGTGACAGC 300

GACAGTTCAA CGCTGCCCCG GAAGTCCCCC TTTGTCCGAA ATACTTTGGA AAGACGAACC 360

CTTCGCTATA AGCAGTCATG CAGGTCTTCC CTGGCTGAGC TCATGGCCCG CACCTCCCTG 420

GACTTGGAGC TGGATCTCCA GGCGTCGAGA ACACGGCAGA GGCAGCTGAA TGAGGAGCTC 480

TGCGCCCTCC GTGAGCTGCG GCAGCGGTTN GGAGGACGCC CAGCTCCGTG GCCAGACTGA 540

CCTCCCACCC TGGGTGCTTC GGGACGAGCG GCTCCGTGGC CTGCTGCGGG AGCCGAGCGG 600

CAGACAAGAC AGACCAAACT TGACTACCGT CATGAGCAGG CGGCTGAGAA GATGCTGAAG 660

AAGGCCTCCA AGGAGATCTA CCAGCTGCGT GGCAGAGCCA CAAAGAGCCC ATCCAAGTGC 720

AGACCTTTAG GGAGAAGATA GCATTCTTCA CAAGGCCAAG GATCAACATA CCTCCTCTCC 780

CAGCCGACGA CGTCTGATGG AGTGCATTGT GCACATGAAG TATTTATCCA CCTGTTTTAT 840

TTTCATGAAG TTCTTAGACT AGCTGAATTT GTCTTTAAAA TATTTGTGCA AAGCTATTAA 900

TATACACATT TTGTAAAAAA AAAAAAAAAA AAACT 935

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

GCAGGGGCGA GGGGYTGGGG ACCGCGGGGC GGACGGGAGC GAGTATGTCC GCTCTGACTC 60

GGCTGGCGTC TTTCGCTCGC GTTGGAGGCC GCCTTTTCAG AAGCGGCTGC GCACGGACTG 120

CTGGAGATGG TGGAGTCCGT CATGCCGGTG GTGGTGTGCA CATTGAGCCC CGGTATAGAC 180

AGTTCCCCCA GCTGACCAGA TCCCAGGTGT TCCAGAGCGA GTTCTTCAGC GGACTCATGT 240

GGTTCTGGAT TCTCTGGCGC TTTTGGCATG ACTCAGAAGA GGTGCTGGGT CACTTTCCGT 300

ATCCTGATCC TTCCCAGTGG ACAGATGAAG AATTAGGTAT CCCTCCTGAT GATGAAGACT 360

GAAGGTGTAG ACTCAGCCTC ACTCTGTACA AGAGCCAGGT GAGAATTTCA AGGATTATCG 420

ACTTCATATT GCACATTAAA GTTACAAATT AAAGTGGCTT GGTCAAGAAT GARAAAAAAA 480

AAAAAAAATT GGGGGGGGGC CCCN 504

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 620 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

```
GAATTCGGCA CGAGGAGGAG GGGAGGCGGG GTAAGTTTGG TGGGAAACTC TGTAATTTCC    60
WTTTTTACTT TCACAGCAAT AGTGCAGAAT CCAGAATGGA TGTCCTCTTT GTAGCCATCT   120
TTGCTGTGCC ACTTATCCTG GGACAAGAAT ATGAGGATGA AGAAAGACTG GGAGAGGATG   180
AATATTATCA GGTGGTCTAT TATTATACAG TCACCCCCAG TTATGATGAC TTTAGTGCAG   240
ATTTCACCAT TGATTACTCC ATATTTGAGT CAGAGGACAG GCTGAACAGG TTGGATAAGG   300
ACATAACAGA AGCAATAGAG ACTACCATTA GTCTTGAAAC AGCACGTGCA GACCATCCGA   360
AGCCTGTAAC TGTGAAACCA GTAACAACGG AACCTCAGAG TCCAGATCTG AACGATGCCG   420
TGTCCAGTTT GCGAAGTCCT ATTCCCCTCC TCCTGTCGTG TGCCTTTGTT CAGGTGGGGA   480
TGTATTTCAT GTAGAAGGTG GAAGAAGGCT GCTATGACTC TTTGGATGGG AGTCTGGCAA   540
GAGGAAATTG GAAGATAAAA TAAATAATAA GTGAAATAAA AAAAAAAAAA AAAAACTCGA   600
GGGGGGGCCC GGTACCCAAT                                               620
```

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 581 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

```
ACAAGGTGTG TGTAAAGTTT ATGTTTGTAA ACTGAATTCT ATCTTAAATC CAAAAAGAAC    60
TCGGGAGTAA TTCATTTTTG TAGCATAAAG ATCCCTAAGT TTTATTTTGA AATATCTGAT   120
TTTTACACGT TAAAAAATAA CAGGGCATCG AGAGGATTCC TAGGTGACAT CCAGACTCCT   180
TTAGCTTTGT GTGTGTGGCA CCGGTTAGTC TGCTTCTCTC TCCTTTCTTG CACTGCTTCA   240
CACAGCCATG CCCTGCCAGC CCGGGCAGGT GCCTTCCTGT CAATGTACAT TTGGGCTTCT   300
GCTCATGCTG CCCTCCCTCC CCTCCCCTGC CTCCCAACCC CGCCCCTTTT GTTCCTCCAT   360
GGAGTACTTC CATGGGTGTG CCTCCCCCAG CCAAGCCATA ATAGGTGGTT TCCCCTTCGC   420
TTCTGTAGCC CTTGCAGACA TCCTCTGTTT ACAGTAGGTG TTGACTTACT TCCCCTCTCC   480
```

CCGSTAAAGC CATAAACTCC TTAAGGACAG GTAGCATTCT TAGTATCTTC GTTCTTCTCA 540

ATGACCAGTA GACCATTAAA CATGTAGCAA ACAAATGTGA A 581

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1843 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

AAACCCAACN CCCTCCGGTC CCCNAAAGAA AGCCCAGCCC AAATCCCAAG CCGGCAGTGA 60

GCCCGCGAAC AAGGCCCTCA AGACGCCCAG NCGAACAAGC AGCCCCCAGG AGGCCCCGCA 120

AGAGAACTCC CTGGCGGCCC AAGCGGGCAG CTTCTGTGCG GCAGAACTCA GCCACCGAGA 180

GCGCAGACAG CATCGAGATT TATGTCCCGG AGNCCCAGAC CAGGCTCTGA GACCATGCAG 240

GAGGAAAGAA ACGATTTTAA ATCATTAAAA ACACAAAAAC TAAGTGCGAA CGGAACAGAG 300

TTTTCTCAAC CTTTGCTATG GTTATTCTGT CTAGAGACCC TGAGCCAACT TTCAAATTGA 360

CGCATACAAG GGCTCACAAT TTGGCTTTTT TGGGTCCCTC CCAGCTTTAG GTTATGAAGA 420

TTTTACTCAC AAAAAAAATC AACAAAAATC ACGAAACTAG AAAACTTTTT TTTTCCTCTT 480

GCTGGCCGTG GTGGACTAGA TAGATGGACG TCGGCAACTC CCGGCCCAGC CTCCATACTG 540

CGGTCTTTTT ACTCGTTCTA TCTGATGAGA ACTCACACTA GCTTGTTTAC AAGATGACGA 600

CAGTCCAAGG GCAGCCTTGG GCACCTGCCA TGTCCCTCCT TTCCCCAGCT ATCCCCGCTC 660

TGACCTTGAT TTTCATTCTT ATGTTTTTCT CTTTTCCCTT CAGAGCTCAC ACAGTGGTCA 720

CCATTGTGGC AAGCGGCTTT CTGGGTCTCA GCCCTCTCTG CGGTTGAGGG CCCAGAGGAC 780

AGAGAGATGG ACATGCGTCC CCTCCCTCCC CCCGCCAAGT GCTCACACAC AACCTCACGC 840

GCACACACAC ACACGCAGAT GGAGGCGCCT CACTGGGAGG TGCCCCGCCA GCCCTGGGCA 900

GTGTCAGGCA GGACTCACTC ACCGCTGAGC AGATGAGAGA AGTTTTAGTC TTGGCGGGTG 960

GAAATGAGAC GAAGCCACAG TTATCACACT CCAGACTCCT GCCCTTTTAT TTTCTCCAGC 1020

CCCTTCTTCC TTCAGCAAAA TCTAGGACTC CCGAGTGGCT TCCAGGGGGC CGTCAGTCCT 1080

CAGCCGCGCC TGTGTCCGGT GCCCGAGGGG CGGGCGGCGG TGTCTGTATG TATGTGTACA 1140

TATGCACATA GACCTTAGAG TGTATAGTTA ACAAACGCCC ATCTGCTCAC CCATGCCCAC 1200

CCAGCGCCGC CGCCGCTGGC TCTCGGGGCA CCTGGCAGGA GGCGGGTGTG TGAATAGCAT 1260

ATATTTTTAC ATGTACTATA TCTAGGTGTG TGTACAAGTG TGTGTAAAAA TATATACCTT 1320

```
GTGTGTAAGC AGCCCTTTTT TTTTTTGGTC TCCACCCCCC TCCCCCCGCC CCGCACTCCT   1380
AAGGGCCCAT CTGCCCAGCC TCTGAGTTTT CTGTTCTATT TTTTTTTTAA CCCCAATTAT   1440
CCTTCTCTCT CTCCTGCCCC CGCATCCCAC TCCCAGGGTG TCACGAGCCC TGAGCTGCAA   1500
TGGCCCGGGC CTGCAGGGCG GGGTAGGGGA GGGCARGGCT SAGCCCCGAA GCCAGCTCAG   1560
TACCTGAGGG GCTGCTCTAT GCTGTGTATG CGCCTCTCTG GCATCCGAGA CATCCTCTTG   1620
GTGGCGCTTG CTNGCAGGGG ACCCCCCCCC CGTCCCCAGG TGAACCAAGG GTCTGCTCCG   1680
GGGCCCATTT CCAGCTTGGC CGCCGTCTGT GACCTTGGGC AAGTCACTTG ACCTCTGTGT   1740
GCCTCAACTT CCTCCTCTGT AAAACGGGGA CAGTCCCTGC CCCTCCCTAC CTCACAGGCA   1800
TGTTGTGAGA ATAAATGAGG TAACGTGTAA AAAAAAAAAA AAT                     1843
```

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

```
TCGACCCACG CGTCCGGCTC CCCGAGCCCT GCCAACCATG GTGAACTTGG GTCTGTCCCG     60
GGTGGACGAC GCCGTGGCTG CCAAGCACCC GGGACTCGGG GAGTATGCCG CATGCCAGTC    120
ACACGCCTTC ATGAAGGGCG TTTTCACCTT CGTCACAGGC ACCGGCATGG CCTTTGGCTT    180
GCAGATGTTC ATTCAGAGGA AGTTTCCATA CCCTTTGCAG TGGAGCCTCC TAGTGGCCGT    240
GGTTGCAGGC TCTGTGGTCA GCTACGGGGT GACGAGAGTG GAGTCGGAGA AATGCAACAA    300
CCTCTGGCTC TTCCTGGAGA CCGGGCAGCT CCCCAAAGAC AGGAGCACAG ATCAGAGAAG    360
CTAGGAGAGC TCCAGCAGGG GCACAGAGGA TTGGGGGCAG GAGGAGTCTG GAACACAGCC    420
TTCATGCCCC CTGACCCCAG GCCGACCCTC CCCACACCCT AGGGTACCCC AGTCGTATCC    480
TCTGTCCGCA TGTKTGGCCA GGCCTGACAA ACACCTGCAG ATGGCTGCTG CCCCAACCTG    540
GGACCTGCCC AGRAGGTTGG AGCAGAAAGG GCTCTCCCTG GGGTGGTGTT TCTCCTCTAG    600
GGTATTGGGA TGCATGTTCT GCACTGCCAG CAGAGAGGGT GTGTCTGGGG GCCACCACCT    660
ATGGGACACG GGGTCGAAGG GGCCTGTACA CTCTGTCATT TCCTTTCTAG CCCCTGCATC    720
TCCAACAAGT CCAAGGTGAC AGCTGGTGCT AGGGGCGTGG GGTTAATAAA TGGCTTATCC    780
TTCTCTCCAC CCAAGTTTCC ACCTGACCAG GTGAAAAACA AATCAGAAGG GTAAGATGAT    840
```

GACAGGTCAC ATGAAACCTT TATTACCCTA CAGTTGATAT ATGAGGATCA CATGCAAGTT 900

ACATACTGAG GATGTACAGG GAAGTTCCCA GCGCTGAACC CCAGAATTAG ACGTTCGCAT 960

CAGCCCCGTA GGCCACGTGG ACACCACCAC AGCCTCTCTG TATGGGGGTC TGCCTCTGTA 1020

GCACTTGGCA TGTAGGGGCA GAGCAAAAGG GGCCANGCTG GCCAGAGCCT GGCTGCTGGG 1080

NAGARGAGGG ACTTGTGGGS CACGCCACNT GCCTATCATT CCCCAYTCAT CTATTAGCCA 1140

AAGTCACTCC CCAGAGGCAG AGCTAGCCCG TTGTAGCCGT GTCTGTGTGG AGGGAAAGCT 1200

TCTGAGTGGG CAAGCCTACA CACAGCCCCG AGCCCCAAGA GGAGGAAGAG GTGGAGACCA 1260

GACGGAACCT CCACAAGTCC ATCATGGTTA CAGCTGGCTT CCCCGCAGCA CCGAAGACCC 1320

ACAGCATNGG CCCTGCTGCC CCCGACCCAG CTCAGCTGCC ANGCCTCACC TTGCCAGGAA 1380

TTGAAAGAAA GTTATTGAGT ACTAATTGGC CTCAGAGTNA CAGGAAGCTC AAGTTAAAGT 1440

G 1441

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 910 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

GGCAGAGCTG GCCTTCGACT CGCTATGTCC ACTAACAATA TGTCGGACCC ACGGAGGCCG 60

AACAAAGTGC TGAGGTGAGG ACCCCAGCGT CGTGGGCACG GGTTCGGGTT GTGGGTGTGG 120

ATCGGGGCCC TGGGAAGCGC CTGTCTATCC CGGGGGCAGG ACCTGAGCGC CCCTGACCCT 180

CGAGCCTGTC GCAGGTACAA GCCCCCGCCG AGCGAATGTA ACCCGGCCTT GGACGACCCG 240

ACGCCGGACT ACATGAACCT GCTGGGCATG ATCTTCAGCA TGTGCGGCCT CATGCTTAAG 300

CTGAAGTGGT GTGCTTGGGT CGCTGTCTAC TGCTCCTTCA TCAGCTTTGC CAACTCTCGG 360

AGCTCGGAGG ACACGAAGCA AATGATGAGT AGCTTCATGT GAGACTTGCC CTACAGAACA 420

AGTGACTCTT GAGTAAGGGG TGGGGGGACC CCAGCCTGGC CATCCTAGAC TGACACCTCT 480

CTCCTGTCTT CATGCTGTCC ATCTCTGCCG TGGTGATGTC CTATCTGCAG AATCCTCAGC 540

CCATGACGCC CCCATGGTGA TACCAGCCTA GAAGGGTCAC ATTTTGGACC CTGTCTATCC 600

ACTAGGCCTG GGCTTTGGCT GCTAAACCTG CTGCCTTCAG CTGCCATCCT GGACTTCCCT 660

GAATGAGGCC GTCTCGGTGC CCCCAGCTGG ATAGAGGGAA CCTGGCCCTT TCCTAGGGAA 720

CACCCTAGGC TTACCCCTCC TGCCTCCCTT CCCCTGCCTG CTGCTGGGGG AGATGCTGTC 780

CATGTTTCTA GGGGTATTCA TTTGCTTTCT CGTTGAAACC TGTTGTTAAT AAAGTTTTTC 840

ACTCTGAAAA AAAAAAAAAA AAAAAAAAAC TYGRGGGGGG GCCCGGAACC CAATTCSCCG 900

GATAGTGAGT 910

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

TCGACCCACG CGTCCGGGCG GGCAGTGATG GCGGCTGGTG ATGGGGACGT GAAGCTAGGC 60

ACCCTGGGGA GTGGCAGCGA GAGCAGCAAC GACGGCGGCA GCGAGAGTCC AGGCGACGCG 120

GGAGCGGCAG CGRAAGGGGG AGGCTGGGCG GCGGCGGCGT TGGCGCTTCT GACGGGGGGC 180

GGGGAAATGC TGCTGAACGT GGCGCTGGTG GCTCTGGTGC TGCTGGGGGC CTACCGGCTG 240

TGGGTGCGCT GGGGGCGGCG GGGTCTGGGG GCCGGGGCCG GGGCGGGCGA GGAGAGCCCC 300

GCCACCTCTC TGCCTCGCAT GAAGAAGCGG GACTTCAGCT TGGAGCAGCT GCGCCAGTAC 360

GACGGCTCCC GCAACCCGCG CATCCTGCTC GCGGTCAATG GGAAAGTCTT CGACGTGACC 420

AAAGGCAGCA AGTTCTACGG CCCGGCGGGT CCATATGGAA TATTTGCTGG TAGGGATGCC 480

TCCAGAGGAC TGGCCACATT TTGCCTAGAT AAAGATGCAC TTAGAGATGA ATATGATGAT 540

CTCTCAGATT TGAATGCAGT ACAAATGGAG AGTGTTCGAG AATGGGAAAT GCAGTTTAAA 600

GAAAAATATG ATTATGTAGG CAGACTCCTA AAACCAGGAG AAGAACCATC AGAATATACA 660

GATGAAGAAG ATACCAAGGA TCACAATAAA CAGGATTGAA CTTTGTAAAC AACCAAAGTC 720

AGGGGCCTTC AGAACTGCAA TTCTTACTCC CTTTCACAGA CTGTCCGGAG TCTTTGGGTT 780

TGATTCACCT GCTGCGAAAA ACATTCAACA AATTGTGTAC AAGATAAATT AATCTCACTA 840

TGAAGATTTG AATAACTAGA CATTATTTAT GCTGCCAAAC TCATTTGTTG CAGTTGTTTG 900

TAATGTCTAG TGGGGCTTCA TCATCCTGAA AAGAAGGAGA CAGGGATTTT TTTAAAGAGC 960

AAGAAAGTCA CAATATTACT TCTTTCCTTC CTTTTTTCCT TCTTTCCTTT CTTCTTTCTC 1020

TTTCTTTCTT TTTAAAATAT ATTGAAGACA ACCAGATATG TATTTGCTAC TCAAGTGTAC 1080

AGATCTCCTC AAGAAACATC AAGGGACTCC TGTGTCACAT ACTGTGTTTT TATTTTAACA 1140

TGGGTGAGGG AGGCGACCTG ATCAGGGGAG GTGGGGGTAC ACATCAATTT GAGTTGTTCA 1200

GGCTACTGAA ACATTAAAAT GTGAATTCCC AAACTTTTCT TTTTGGCTTT GTCAGGGAAA 1260

AGAAAAATAT CTTTATAAAG AAATCTTTGG AAATTAGGAG AAGGAATTTC AGGTGGGTTT 1320

AAGTCAGAGC TAGTTCCCCA ACAGAAAGAT CATTTGAAAC CAGTTTTTAT CCCTTCTCTT 1380

TCCTTCCCTT TCCCTAAATC AAATCAATAT TAATTGTGCC TTATTTCACT TAACATAGAC 1440

TTGAATTATT TTTAGGGAAA GCCCCTATAA TGAATTCAGA AATCACTACA AGCAGCATTA 1500

AGACTGAAGT TGGAATATTC TGTTGACCAT AAAACCTTGA TATCATTCTG TGTATATAGA 1560

ATGTAAAAGG AATATTACAG TGTTAACTGC CATATATGTA ATATACACAA ACTCAATTAG 1620

CATTGTAATG GCCAAATGCA TTCCCCCATG CTTTTCTGTT TTCAAAAAAA TTGAAAAACA 1680

AATCAACTCT TATCCCCAAC AGCTGCCTAA TTTTAGGAGT CTGACCCTCC ACATCTCACT 1740

GGTGTGGGTG CATGGGGCTG TGGAGTGGGT GTCAGTATGG ATGTGTCTGA ATGTGTGAGG 1800

CCTTGGAAGG GACTCTTTCT GCAGATACTG TAAATACAAG TACCATTTTA ATAAAGCATG 1860

TACAATAAAC CAAAATAAGC TTGAGTTGGA CTTTATATAC AGAACTGTAA GCCAGTGCAT 1920

TATGATACAG TTGTAAGATT GTGCATTTGA TTCAAGATAA GGAAAAATCT TGGAAATGAA 1980

AAGCAGGCAC KGGTTAACCA AGTTGTACAC ATTGTACCAC ATTCAGCATA ACTTTAGGAA 2040

GAAATTCCAC TTTGTGAACA TTCTCCAGAA ATCCAAGATT ATTCAGGTAA GAATTGGTAT 2100

ATTAAATGTA CATCTTTTTA CTTTCTATTT TGATGCCAAC TGATTATACT AGACAATTAG 2160

CACTCCAGGT GGTTATTGAA CACAAAACAG TAAAAGAATA TTGCACTGAT AGATACTAAA 2220

TTATTATTTT ATTAGGTTGA AAAAGCCCTT ACTAAAAGCC CCTCATATAT CAATTACTTT 2280

ATTTCATTAT GACTACTTAG GTTCCGGGCT GGGGACAAGT TCACTTAAAA AGGCAATGTT 2340

ATTTAACAGG TCACCAGTTA AGACTTCTGC TTTGTAGATA CATGCAGAAG CCATCAAACA 2400

AGGGGGRGCT TTTAACTGCA ACAATAAGCT AAAGTATGTA AAATACTACA TTCTATTCAG 2460

TCTTGGAGTG TTTTGTAGAA AGTTATCTTC AGCCAAATCT TTGCTGAAGA CTGGTTGTGG 2520

AGTGTTGGTA AATGCTTTGT GTTTTTATGT AAAATATTTT CTAAACAAAA AATGTTAAAA 2580

GTACATGTCC TCTGTAGTAA ACTGATATCT ATATATATGA ATCATTCAAG CCTAAAGTCT 2640

AGTAATAAAC TGTACTTGTG AATAGAGAAA CCCTAAATAT TCATGCAGWA AAAATTATGC 2700

GGTCTGTTAA GAAAAATGAG TAATTTGTGT TTTGGACTTG AAATAAACAG TGTTCTGTAG 2760

ATAATTCCTC AACTTC 2776

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1525 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

```
CCGCTGCTGA TAACTATGGC ATCCCCCGGG CCTGCAGGAA TTCGGCACGG AGCTACGGCG      60
CCGCCTGGCT CCTGCTGNCA CCTGCAGGCT CGTCGCGGGT GGAGCCCACC CAAGACATCA     120
GCATCAGCGA CCAGCTGGGG GGCCAGGACG TGCCCGTGTT CCGGAACCTG TCCCTGCTGG     180
TGGTGGGTGT CGGCGCCGTG TTCTCACTGC TATTCCACCT GGGCACCCGG GAGAGGCGCC     240
GGCCGCATGC GGASGAGCCA GGCGAGCACA CCCCCCTGTT GGCCCCTGCC ACGGCCCAGC     300
CCCTGCTGCT CTGGAAGCAC TGGCTCCGGG AGCSGGCTTT CTACCAGGTG GGCATACTGT     360
ACATGACCAC CAGGCTCATC GTGAACCTGT CCCAGACCTA CATGGCCATG TACCTCACCT     420
ACTCGCTCCA CCTGCCCAAG AAGTTCATCG CGACCATTCC CCTGGTGATG TACCTCAGCG     480
GCTTCTTGTC CTCCTTCCTC ATGAAGCCCA TCAACAAGTG CATTGGGAGG AACATGACCT     540
ACTTCTCAGG CCTCCTGGTG ATCCTGGCCT TTGCCGCCTG GGTGGCGCTG GCGGAGGGAC     600
TGGGTGTGGC CGTGTACGCA GCGGCTGTGC TGCTGGGTGC TGGCTGTGCC ACCATCCTCG     660
TCACCTCGCT GGCCATGACG GCCGACCTCA TCGGTCCCCA CACGAACAGC GGACTKTCGT     720
GTACGGCTCC ATGAGCTTCT TGGATAAGGT GGCCAATGGG CTGGCAGTCA TGGCCATCCA     780
GAGCCTGCAC CCTTGCCCCT CAGAGCTCTG CTGCAGGGCC TGCGTGAGCT TTTACCACTG     840
GGCGATGGTG GCTGTGACGG GCGGCGTGGG CGTGGCCGCT GCCCTGTGTC TCTGTAGCCT     900
CCTGCTGTGG CCGACCCGCC TGCGACGCTG GGACCGTGAT GCCCGGCCCT GACTCCTGAC     960
AGCCTCCTGC ACCTGTGCAA GGGAACTGTG GGGACGCACG AGGATGCCCC CCARGGCCTT    1020
GGGGAAAAGC CCCCACTGCC CCTCACTCTT CTCTGGACCC CCACCCTCCA TCCTCACCCA    1080
GCTCCCGGGG GTGGGGTCGG GTGAGGGCAG CAGGGATGCC CGCCAGGGAC TTGCAAGGAC    1140
CCCCTGGGTT TTGAGGGTGT CCCATTCTCA ACTCTAATCC ATCCCAGCCC TCTGGAGGAT    1200
TTGGGGTGCC CCTCTCGGCA GGGAACAGGA AGTAGGAATC CCAGAAGGGT CTGGGGGAAC    1260
CCTAACCCTG AGCTCAGTCC AGTTCACCCC TCACCTCCAG CCTGGGGGTC TCCAGACACT    1320
GCCAGGGCCC CCTCAGGACG GCTGGAGCCT GGAGGAGACA GCCACGGGGT GGTGGGCTGG    1380
GCCTGGACCC CACCGTGGTG GGCAGCAGGG CTGCCCGGCA GGCTTGGTGG ACTCTGCTGG    1440
CAGCAAATAA AGAGATGACG GCAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA    1500
AAAAAAAAAA AAACCCACCG TCCGC                                          1525
```

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1563 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

```
AATTCGGCAC GAGNCAGAAA CCTGCGGAAA ATGGTAGCGA TGGCGGCTGG GCCGAGTGGG      60
TGTCTGGTGC CGGCGTTTGG GCTACGGTTG TTGTTGGCGA CTGTGCTTCA AGCGGTGTCT     120
GCTTTTGGGG CAGAGTTTTC ATCGGAGGCA TGCAGAGAGT TAGGCTTTTC TAGCAACTTG     180
CTTTGCAGCT CTTGTGATCT TCTCGGACAG TTCAACCTGC TTCAGCTGGA TCCTGATTGC     240
AGAGGATGCT GTCAGGAGGA AGCACAATTT GAAACCAAAA AGCTGTATGC AGGAGCTATT     300
CTTGAAGTTT GTGGATGAAA ATTGGGAAGG TTCCCTCAAG TCCAAGCTTT TGTTAGGAGT     360
GATAAACCCA AACTGTTCAG AGGACTGCAA ATCAAGTATG TCCGTGGTTC AGACCCTGTA     420
TTAAAGCTTT TGGACGACAA TGGGAACATT GCTGAAGAAC TGAGCATTCT CAAATGGAAC     480
ACAGACAGTG TAGAAGAATT CCTGAGTGAA AAGTTGGAAC GCATATAAAT CTTGCTTAAA     540
TTTTGTCCTA TCCTTTTGTT ACCTTATCAA ATGAAATATT ACAGCACCTA GAAAATAATT     600
TAGTTTTGCT TGCTTCCATT GATCAGTCTT TTACTTGAGG CATTAAATAT CTAATTAAAT     660
CGTGAAATGG CAGTATAGTC CATGATATCT AAGGAGTTGG CAAGCTTAAC AAAACCCATT     720
TTTTATAAAT GTCCATCCTC CTGCATTTGT TGATACCACT AACAAAATGC TTTGTAACAG     780
ACTTGCGGTT AATTATGCAA ATGATAGTTT GTGATAATTG GTCCAGTTTT ACGAACAACA     840
GATTTCTAAA TTAGAGAGGT TAACAAGACA GATGATTACT ATGCCTCATG TGCTGTGTGC     900
TCTTTGAAAG GAATGACAGC AGACTACAAA GCAAATAAGA TATACTGAGC CTCAACAGAT     960
TGCCTGCTCC TCAGAGTCTC TCCTATTTTT GTATTACCCA GCTTTCTTTT TAATACAAAT    1020
GTTATTTATA GTTTACAATG AATGCACTGC ATAAAAACTT TGTAGCTTCA TTATTGTAAA    1080
ACATATTCAA GATCCTACAG TAAGAGTGAA ACATTCACAA AGATTTGCGT TAATGAAGAC    1140
TACACAGAAA ACCTTTCTAG GGATTTGTGT GGATCAGATA CATACTTGGC AAATTTTTGA    1200
GTTTTACATT CTTACAGAAA AGTCCATTTA AAAGTGATCA TTTGTAAGAC CAAAATATAA    1260
ATAAAAAGTT TCAAAAATCT ATCTGAATTT GGAATTCTTC TGGTTTGTTC TTTCATGTTT    1320
AAAAATGATG TTTTTCAATG CATTTTTTTC ATGTAAGCCC TTTTTTTAGC CAAAATGTAA    1380
AAATGGCTGT AATATTTAAA ACTTATAACA TCTTATTGTT GGTAATAGTG CTTTATATTT    1440
```

GTCTGATTTT ATTTTTCAAA GTTTTTTCAT TTATGAACAC ATTTTCATTG GTATATTATT 1500

TAAGGAATAT CTCTTGATAT AGAATTTTTA TATTAAAAAT GATTTTTCTT TGCTTAAAAA 1560

AAA 1563

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

TGCACGCTGG CCATGTGGGN GTTGGGCCAC TGCGACCCCC GGCGCTGCAC GGGCCGCAAG 60

CTGGCCCGCC TGGGGCTGGT GCGCTGCCTG CGCCTGGGCC ACAGATTCGG CGGTCTGGTG 120

CTGAGCCCCG TGGGCAAGCA GTACGCGTCC CCCGCAGACA GACAGCTGGT GGCGCAGTCT 180

GGGGTCGCCG TCATCGACTG CTCCTGGGCC AGGCTGGACG AGACACCGTT TGGGAAGATG 240

CGAGGGAGCC ACTTGCGCCT GTTGCCCTAC CTGGTGGCCG CCAACCCCGT GAACTATGGC 300

CGGCCCTACA GACTTTCCTG CGTGGAAGCG TTTGCTGCCA CCTTCTGCAT CGTAGGCTTT 360

CCAGACCTTG CTGTCATTTT GCTGCGGAAG TTTAAATGGG GCAAGGGCTT CTTGGACCTG 420

AACCGCCAGC TCCTGGACAA GTACGCGGCC TGCGGCAGCC CGGAGGAGGT GCTGCAGGCG 480

GAGCAGGAGT TCTTGGCCAA TGCCAAGGAG AGCCCCCAGG AGGAGGAGAT CGATCCCTTC 540

GATGTGGATT CAGGGAGAGA GTTTGGAAAC CCCAACAGGC CTGTGGCCAG CACCCGGCTG 600

CCCTCGGACA CTGATGACAG TGATGCGTCT GAGGACCCAG GGCCTKGCGC CGAGCGCGGA 660

GGAGCCAGCA GCAGCTGCTG TGAAGAGGAG CAGACGCAGG GACGGGGGGC TGAGGCCAGG 720

GCCCCGGCTG AGGTTTGGAA AGGAATCAAG AAACGGCAGA GAGACTGAGG GTTGCAGACA 780

CATATATTTT TGAGGCTGGG TGACGAGAAA ATCTAGAGAC ATGAGGGACA TAAATGGGCC 840

TGGCAGCCTC GGCTCTTTGC GGCTGCTGGC AGGACTGAGC TGTCCGGGTT CTCCCCACAC 900

TTCCAGCACA GCTGTGCTCT GTGTCCTGCC TCGGCGCTCT CGCAAATGAA GCTGCAGGCC 960

AAGAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAG GGGGGGGGGC 1020

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 770 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

```
TCGACCCACG CGTCCGGGCC GCCGTAGCGC GTCTTGGGTC TCCCGGCTGC CGCTGCTGCC     60
GCCGCCGCCT CGGGTCGTGG AGCCAGGAGC GACGTCACCG CCATGGCAGG CATCAAAGCT    120
TTGATTAGTT TGTCCTTTGG AGGAGCAATC GGACTGATGT TTTTRATGCT TGGATGTGCC    180
CTTCCAATAT ACAACAAATA CTGGCCCCTC TTTGTTCTAT TTTTTTACAT CCTTTCACCT    240
ATTCCATACT GCATAGCAAG AAGATTAGTG GATGATACAG ATGCTATGAG TAACGCTTGT    300
AAGGAACTTG CCATCTTTCT TACAACGGGC ATTGTCGTGT CAGCTTTTGG ACTCCCTATT    360
GTATTTGCCA GAGCACATCT GATTGAGTGG GGAGCTTGTG CACTTGTTCT CACAGGAAAC    420
ACAGTCATCT TTGCAACTAT ACTAGGCTTT TTCTTGGTCT TTGGAAGCAA TGACGACTTC    480
AGCTGGCAGC AGTGGTGAAA AGAAATTACT GAACTATTGT CAAATGGACT TCCTGTCATT    540
TGTTGGCCAT TCACGCACAC AGGAGATGGG GCAGTTAATG CTGAATGGTA TAGCAAGCCT    600
CTTGGGGGTA TTTTAGGTGC TCCCTTCTCA CTTTTATTGT AAGCATACTA TTTTCACAGA    660
GACTTGCTGA AGGATTAAAA GGATTTTCTC TTTTGGAAAA AAAAAAAAAA AAAAACYCGA    720
GGGGGGGCCC GTWCCCATTC SCCCYATATG AATTCCNTTT TTACAATCCC               770
```

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 481 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

```
GAATTCGGCA CGAGCATAGT GTTAACCACT AGAATTCACT GCCCTTCCTA TCCAAAAATG     60
ACACTACTGA TCATTTTTCT TCCTTTTSCT TTTACAACAT TMACAAATTC AGGTGGCTCT    120
TTCCCAGTAC GGTAGGCTGA TTCGTATGGA TGCACCACGG TTGGTGACTC CCCCCACCCC    180
ACAGAGTTTC TGGCGTTCAT TCGGTTGAAC CCAAGGCCAG CAAGGGCTGA CTGGGAACAA    240
ACCGAACACT AGGCCGTGAA CCAATCGTCT CTCCGTGCCC GGGAGCGAMC CCGGGGGCCT    300
TTCACTCTCC CAAGGACTCC ANGGGGGGGC CGGGTACCCA ATTCCGCCCC TATAGTGAAT    360
CCGTNATTAC AATTCCACNT GGGCCGTCCN TTTTTACAAA CGTTCCGTTG AACTGGGAAA    420
AACCCCTTGG CGGTTTACCC CAACTTTAAT CCGCCTTTGC AAGCACATCC CCCCCCTTTT    480
```

C 481

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 644 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

GCTGGGATAG AGCATGAAAG GAGAACTGCT CCCTTTTCTG TTTCTCACAG TTTGGTTATG 60
GCTTTATAAA CTTKTATTTG GTGAAAGCCC CAGATACCCA AATGTCATTG GCAAAACTTA 120
TTTTTTTTTC TGGACAGATC AGATTTCTAG AGAGAGCAGA TTTCTAGAGA GATTAGCATT 180
CATAGTAAGT GAAAATTGTC TAATTTTTTT AATCCATGCT ATTACTGGGC AGTAGGTCTA 240
ATTTTTTTTG ACAAAAAATA GATCTATTTT CCTTATATAT TGATTTAGAA TCTTAAGTTA 300
GAATTTTATA GAAGAAATGT CTGAGCAGTT CTATGTATGG AGGAGCAATT CAGCTTTTCA 360
GCAGCAACTT TATCTTTTGC CACTAGAGGG AGATCTGTGG TTGCTTTCTC CTTTGGAGAA 420
TAGCTGCTTT GCTTTTATTT TTAATTTCTA AGGTTGGAAT AGAACTTATT CTCAAAATTC 480
CTTTAGTGTT ATTAAATATT TTCATTTATT AGTCAAAGGT AAGTTAATTA AGCTTGTTTA 540
ATGATGCCAA TCTTATGCTT TTCTGTAATC TTCAATTTTT AATAAATGTG AGTTAGATAC 600
TAAGTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA 644

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1351 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

GGCACGAGTG CGCASGCGTG GGGCTCTCTC CTTGTCAGTC GGCGCCGCGT GCGGGCTGGT 60
GGCTCTGTGG CAGCGGCGGC GGCAGGACTC CGGCACTATG AGCGGCTTCA GCACCGAGGA 120
GCGCGCCGCG CCNTTCTCCC TGGAGTACCG AGTCTTCCTC AAAAATGAGA AAGGACAATA 180
TATATCTCCA TTTCATGATA TTCCAATTTA TGCAGATAAG GATGTGTTTC ACATGGTAGT 240
TGAAGTACCA CGCTGGTCTA ATGCAAAAAT GGAGATTGCT ACAAAGGACC CTTTAAACCC 300

TATTAAACAA GATGTGAAAA AAGGAAAACT TCGCTATGTT GCGAATTTGT TCCCGTATAA 360

AGGATATATC TGGAACTATG GTGCCATCCC TCAGACTTGG GAAGACCCAG GGCACAATGA 420

TAAACATACT GGCTGTTGTG GTGACAATGA CCCAATTGAT GTGTGTGAAA TTGGAAGCAA 480

GGTATGTGCA AGAGGTGAAA TAATTGGCGT GAAAGTTCTA GGCATATTGG CTATGATTGA 540

CGAAGGGGAA ACCGACTGGA AAGTCATTGC CATTAATGTG GATGATCCTG ATGCAGCCAA 600

TTATAATGAT ATCAATGATG TCAAACGGCT GAAACCTGGC TACTTAGAAG CTACTGTGGA 660

CTGGTTTAGA AGGTATAAGG TTCCTGATGG AAAACCAGAA AATGAGTTTG CGTTTAATGC 720

AGAATTTAAA GATAAGGACT TTGCCATTGA TATTATTAAA AGCACTCATG ACCATTGGAA 780

AGCATTAGTG ACTAAGAAAA CGAATGGAAA AGGAATCAGT TGCATGAATA CAACTTTGTC 840

TGAGAGCCCC TTCAAGTGTG ATCCTGATGC TGCCAGAGCC ATTGTGGATG CTTTACCACC 900

ACCCTGTGAA TCTGCCTGCA CAGTACCAAC AGACGTGGAT AAGTGGTTCC ATCACCAGAA 960

AAACTAATGA GATTTCTCTG GAATACAAGC TGATATTGCT ACATCGTGTT CATCTGGATG 1020

TATTAGAAGT AAAAGTAGTA GCTTTTCAAA GCTTTAAATT TGTAGAACTC ATCTAACTAA 1080

AGTAAATTCT GCTGTGACTA ATCCAATATA CTCAGAATGT TATCCATCTA AAGCATTTTT 1140

CATATCTCAA CTAAGATAAC TTTTAGCACA TGCTTAAATA TCAAAGCAGT TGTCATTTGG 1200

AAGTCACTTG TGAATAGATG TGCAAGGGGA GCACATATTG GATGTATATG TTACCATATG 1260

TTAGGAAATA AAATTATTTT GCTGAAAAAA AAAAAAAAAA AACCNCGGGG GGGGCCCCGG 1320

TCCCCATTTG GCCCTTTGGG GGGNGGTTTT A 1351

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2527 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

CTCTTGCTAC CTTCCCGGCG CAGAGAACCC CGGCTGCTCA GCGCGCTCCG GGGTCATGGA 60

GATCCCCGGG AGCCTGTGCA AGAAAGTCAA GCTGAGCAAT AACGCGCAGA ACTGGGGAAT 120

GCAGAGAGCA ACCAATGTCA CCTACCAAGC CCATCATGTC AGCAGGAACA AGAGAGGTCA 180

GGTGGTGGGG ACCAGAGGTG GCTTTCGTGG TTGCACAGTT TGGCTAACAG GCTTGTCTGG 240

AGCGGGAAAG ACTACTGTGA GCATGGCCTT GGAGGAGTAC CTGGTTTGTC ATGGTATTCC 300

ATGCTACACT CTGGATGGTG ACAATATTCG TCAAGGTCTC AATAAAAATC TTGGCTTTAG 360

TCCTGAAGAC AGAGAAGAGA ATGTTCGACG CATCGCAGAA GTTGCTAAAC TGTTTGCAGA 420

TGCTGGCTTA GTGTGCATCA CAAGTTTCAT ATCACCTTAC ACTCAGGATC GCAACAATGC 480

AAGGCAAATT CATGAAGGTG CAAGTTTACC GTTTTTTGAA GTATTTGTTG ATGCTCCTCT 540

GCATGTTTGT GAACAGAGGG ATGTCAAAGG ACTCTACAAA AAAGCCCGGG CAGGAGAAAT 600

TAAAGGTTTC ACTGGGATCG ATTCTGAATA TGAAAAGCCA GAGGCCCCTG AGTTGGTGCT 660

GAAAACAGAC TCCTGTGATG TAAATGACTG TGTCCAGCAA GTTGTGGAAC TTCTACAGGA 720

ACGGGATATT GTACCTGTGG ATGCATCTTA TGAAGTAAAA GAACTATATG TGCCAGAAAA 780

TAAACTTCAT TTGGCAAAAA CAGATGCGGA AACATTACCA GCACTGAAAA TTAATAAAGT 840

GGATATGCAG TGGGTGCAGG TTTTGGCAGA AGGTTGGGCA ACCCCATTGA ATGGCTTTAT 900

GAGAGAGAGG GAGTACTTGC AGTGCCTTCA TTTTGATTGT CTTCTGGATG GAGGTGTCAT 960

TAACTTGTCA GTACCTATAG TTCTGACTGC GACTCATGAA GATAAAGAGA GGCTGGACGG 1020

CTGTACAGCA TTTGCTCTGA TGTATGAGGG CCGCCGTGTG GCCATTCTTC GCAATCCAGA 1080

GTTTTTTGAG CACAGGAAAG AGGAGCGCTG TGCCAGACAG TGGGGAACGA CATGCAAGAA 1140

CCACCCCTAT ATTAAGATGG TGATGGAACA AGGAGATTGG CTGATTGGAG GAGATCTTCA 1200

AGTCTTGGAT CGAGTTTATT GGAATGATGG TCTTGATCAG TATCGTCTTA CTCCTACTGA 1260

GCTAAAGCAG AAATTTAAAG ATATGAATGC TGATGCTGTC TTTGCATTTC AACTACGCAA 1320

CCCAGTGCAC AATGGACATG CCCTGTTAAT GCAGGATACC CATAAGCAAC TTCTAGAGAG 1380

GGGCTACCGG CGCCCTGTCC TCCTCCTCCA CCCTCTGGGT GGCTGGACAA AGGATGACGA 1440

TGTTCCTTTG ATGTGGCGTA TGAAGCAGCA TGCTGCAGTG TTGGAGGAAG GAGTTCTGAA 1500

TCCTGAGACG ACAGTGGTGG CCATCTTCCC ATCTCCCATG ATGTATGCTG GACCAACTGA 1560

GGTCCAGTGG CATTGCAGAG CACGGATGGT TGCAGGAGCC AACTTTTACA TTGTTGGACG 1620

AGACCCTGCT GGCATGCCTC ATCCAGAAAC AGGGAAGGAT CTTTATGAGC CAAGTCATGG 1680

TGCCAAAGTG CTGACGATGG CCCCTGGTTT AATCACTTTG GAAATAGTTC CCTTTCGAGT 1740

TGCAGCTTAC AACAAGAAAA AGAAGCGTAT GGACTACTAT GACTCTGAAC ACCATGAAGA 1800

CTTTGAATTT ATTTCAGGAA CACGAATGCG CAAACTTGCT CGAGAAGGCC AGAAACCACC 1860

TGAAGGTTTC ATGGCTCCCA AGGCTTGGAC CGTGCTGACA GAATACTACA AATCCTTGGA 1920

GAAAGCTTAG GCTGTTAACC CAGTCACTCC ACCTTTGACA CATTACTAGT AACAAGAGGG 1980

GACCACATAG TCTCTGTTGG CATTTCTTTG TGGTGTCTGT CTGGACATGC TTCCTAAAAA 2040

CAGACCATTT TCCTTAACTT GCATCAGTTT TGGTCTGCCT TATGAGTTCT GTTTTGAACA 2100

AGTGTAACAC ACTGATGGTT TTAATGTATC TTTTCCACTT ATTATAGTTA TATTCCTACA 2160

ATACAATTTT AAAATTGTCT TTTTATATTA TATTTATGCT TCTGTGTCAT GATTTTTTCA 2220

AGCTGTTATA TTAGTTGTAA CCAGTAGTAT TCACATTAAA TCTTGCTTTT TTTCCCCTTA 2280

AAAAAAGAAA AAAATTACCA AACAATAAAC TTGGCTAGAC CTTGTTTTGA GGATTTTACA 2340

AGACCTTTGT AGCGATTAGA TTTTTTTTCT ACATTGAAAA TAGAAACTGC TTCCTTTCTT 2400

CTTTCCAGTC AGCTATTGGT CTTTCCAGCT GTTATAATCT AAAGTATTCT TATGATCTGT 2460

GTAAGCTCTG AATGAACTTC TTTACTCAAT AAAATTAATT TTTTGGCTTC TTAAAAAAAA 2520

AAAAAAA 2527

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2566 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CCCAAGAATT CGGCACGAGC GNGGCAWAAK TGGGATTTCT GAAACCTGTA GGCCCCAAGC 60

CCATCAACTT GCCCAAAGAA GATTCCAAAC CTACATTTCC CTGGCCTSCT GGAAACAAGC 120

CATCTCTTCA CAGTGTAAAC CAAGACCATG ACTTAAAGCC ACTAGGCCGA AATCTGGGCC 180

TACTCCTCCA ACCTCAGAAA ATGAACAGAA GCAAGCKTTT CCCAAATTGA CTGGGGTTAA 240

AGGGAAATTT ATGTCAGCAT CACAAGATCT TGAACCCAAG CCCCTCTTCC CCAAACCCGC 300

CTTTGGCCAG AAGCCGCCCC TAAGTACCGA GAACTCCCAT GAAGACGAAA GCCCCATGAA 360

GAATGTGTCT TCATCAAAAG GGTCCCCAGC TCCCCTGGGA GTCAGGTCCA AAAGCGGCCC 420

TTTAAAACCA GCAAGGGAAG ACTCAGAAAA TAAAGACCAT GCAGGGGAGA TTTCAAGTTT 480

GCCCTTTCCT GGAGTGGTTT TGAAACCTGC TGCGAGCAGG GGAGGCCCAG GTCTCTCCAA 540

AAATGGTGAA GAAAAAAAGG AAGATAGGAA GATAGATGCT GCTAAGAACA CCTTCCAGAG 600

CAAAATAAAT CAGGAAGAGT TGGCCTCAGG GACTCCTCCT GCCAGGTTCC CTAAGGCCCC 660

TTCTAAGCTG ACAGTGGGGG GGCCATGGGG CCAAAGTCAG GAAAAGGAAA AGGGAGACAA 720

GAATTCAGCC ACCCCGAAAC AGAAGCCATT GCCTCCCTTG TTTACCTTGG GTCCACCTCC 780

ACCAAAACCC AACAGACCAC CAAATGTTGA CCTGACGAAA TTCCACAAAA CCTCTTCTGG 840

AAACAGTACT AGCAAAGGCC AGACGTCTTA CTCAACAACT TCCCTGCCAC CACCTCCACC 900

ATCCCATCCG GCCAGCCAAC CACCATTGCC AGCATCTCAC CCATCACAAC CACCAGTCCC 960

AAGCCTACCT CCCAGAAACA TTAAACCTCC GTTTGACCTA AAAAGCCCTG TCAATGAAGA 1020

CAATCAAGAT GGTGTCACGC ACTCTGATGG TGCTGGAAAT CTAGATGAGG AACAAGACAG 1080

TGAAGGAGAA ACATATGAAG ACATAGAAGC ATCCAAAGAA AGAGAGAAGA AAAGGGAAAA 1140

GGAAGAAAAG AAGAGGTTAG AGCTGGAGAA AAAGGAACAG AAAGAGAAAG AAAAGAAAGA 1200

ACAAGAAATA AAGAAGAAAT TTAAACTAAC AGGCCCTATT CAAGTCATCC ATCTTGCAAA 1260

AGCTTGTTGT GATGTCAAAG GAGGAAAGAA TGAACTGAGC TTCAAGCAAG GAGAGCAAAT 1320

TGAAATCATC CGCATCACAG ACAACCCAGA AGGAAAATGG TTGGGCAGAA CAGCAAGGGG 1380

TTCATATGGC TATATTAAAA CAACTGCTGT AGAGATTGAC TATGATTCTT TGAAACTGAA 1440

AAAAGACTCT CTTGGTGCCC CTTCAAGACC TATTGAAGAT GACCAAGAAG TATATGATGA 1500

TGTTGCAGAG CAGGATGATA TTAGCAGCCA CAGTCAGAGT GGAAGTGGAG GGATATTCCC 1560

TCCACCACCA GATGATGACA TTTATGATGG GATTGAAGAG GAAGATGCTG ATGATGGCTC 1620

CACACTACAG GTTCAAGAGA AGAGTAATAC GTGGTCCTGG GGGATTTTGA AGATGTTAAA 1680

GGGAAAAGAT GACAGAAAGA AAAGTATACG AGAGAAACCT AAAGTCTCTG ACTCAGACAA 1740

TAATGAAGGT TCATCTTTCC CTGCTCCTCC TAAACAATTG GACATGGGAG ATGAAGTTTA 1800

CGATGATGTG GATACCTCTG ATTTCCCTGT TTCATCAGCA GAGATGAGTC AAGGAACTAA 1860

TGTTGGAAAA GCTAAGACAG AAGAAAAGGA CCTTAAGAAG CTAAAAAAGC AGRAAAAARA 1920

ARAAAAAGAC TTCAGGAAAA AATTTAAATA TGATGGTGAA ATTAGAGTCC TATATTCAAC 1980

TAAAGTTACA ACTTCCATAA CTTCTAAAAA GTGGGGAACC AGAGATCTAC AGGTAAAACC 2040

TGGTGAATCT CTAGAAGTTA TACAAACCAC AGATGACACA AAAGTTCTCT GCAGAAATGA 2100

AGAAGGGAAA TATGGTTATG TCCTTCGGAG TTACCTAGCG GACAATGATG GAGAGATCTA 2160

TGATGATATT GCTGATGGCT GCATCTATGA CAATGACTAG CACTCAACTT TGGTCATTCT 2220

GCTGTGTTCA TTAGGTGCCA ATGTGAAGTC TGGATTTTAA TTGGCATGTT ATTGGGTATC 2280

AAGAAAATTA ATGCACAAAA CCACTTATTA TCATTTGTTA TGAAATCCCA ATTATCTTTA 2340

CAAAGTGTTT AAAGTTTGAA CATAGAAAAT AATCTCTCTG CTTAATTGTT ATCTCAGAAG 2400

ACTACATTAG TGAGATGTAA GAATTATTAA ATATTCCATT TCCGCTTTGG CTACAATTAT 2460

GAAGAAGTTG AAGGTACTTC TTTTAGACCA CCAGTAAATA ATCCTCCTTC AAAAAATAAA 2520

AATAAAAAAA AAAAAAAAAA ACTCGAGGGG GGGCCCGGTA CCCAAT 2566

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

```
GAATTCGGCA CGAGGCTTTC TGTGTCCTCT GTGGCTGCTT TAGTGTGCCA CCAGGGGCAG    60
ACTTGGGTGG GTTGCAGCAG AGATGGCATG GCCCTCAAGG TCCAAGATGT TTACTCTCTT   120
GCCGGTCCTC TGTTATCTCT GGTCTTTGTG GTTGCCACAG TTTTCTTGGA TCCAGGAGTT   180
AAAGGCAGTC CTGAGGGATG ATGGCCTCAT CTCCGCAGTT GCYTGGAATG CTGAATTTCA   240
GACGTGCTAA AGGAGGGTTG CAGACATTGT GTGGWATGCA TTCAGACCCC AGATGTGGGT   300
GCAGGAAGGC AGGCATGGCA CAGCCAGGTA GAGACTGGTT TCCAGGCCCA AGCAGCCTTC   360
AGCAGCTGTG CGCCTTGTTT CTGATGTTGT TTGGGAGTAA GAATAATGTA GACATGGGGG   420
GTCATGARGC TCAATAAAAA CTTCAAGGAA ACCTCCCATG GCATGGTTGG GCGCAGTGAC   480
TCATGCCTGT AACCCCAGCA CTGTGGAATG CCAAGGTGGA AGGATCGCTT GAGGCCAAGA   540
```

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1863 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

```
TCGACCCACG CGTCCGGCGA GATCCCTACC GCAGTAGCCG CCTCTGCCGC CGCGGAGCTT    60
CCCGAACCTC TTCAGCCGCC CGGAGCCGCT CCCGGAGCCC GGCCGTAGAG GCTGCAATCG   120
CAGCCGGGAG CCCGCAGCCC GCGCCCCGAG CCCGCCGCCG CCCTTCGAGG GCGCCCCAGG   180
CCGCGCCATG GTGAAGGTGA CGTTCAACTC CGCTCTGGCC CAGAAGGAGG CCAAGAAGGA   240
CGAGCCCAAG AGCGGCGAGG AGGCGCTCAT CATCCCCCCC GACGCCGTCG CGGTGGACTG   300
CAAGGACCCA GATGATGTGG TACCAGTTGG CCAAAGAAGA GCCTGGTGTT GGTGCATGTG   360
CTTTGGACTA GCATTTATGC TTGCAGGTGT TATTCTAGGA GGAGCATACT TGTACAAATA   420
TTTTGCACTT CAACCAGATG ACGTGTACTA CTGTGGAATA AAGTACATCA AAGATGATGT   480
CATCTTAAAT GAGCCCTCTG CAGATGCCCC AGCTGCTCTC TACCAGACAA TTGAAGAAAA   540
TATTAAAATC TTTGAAGAAG AAGAAGTTGA ATTTATCAGT GTGCCTGTCC CAGAGTTTGC   600
AGATAGTGAT CCTGCCAACA TTGTTCATGA CTTTAACAAG AAACTTACAG CCTATTTAGA   660
```

TCTTAACCTG GATAAGTGCT ATGTGATCCC TCTGAACACT TCCATTGTTA TGCCACCCAG 720

AAACCTACTG GAGTTACTTA TTAACATCAA GGCTGGAACC TATTTGCCTC AGTCCTATCT 780

GATTCATGAG CACATGGTTA TTACTGATCG CATTGAAAAC ATTGATCACC TGGGTTTCTT 840

TATTTATCGA CTGTGTCATG ACAAGGAAAC TTACAAACTG CAACGCAGAG AAACTATTAA 900

AGGTATTCAG AAACGTGAAG CCAGCAATTG TTTCGCAATT CGGCATTTTG AAAACAAATT 960

TGCCGTGGAA ACTTTAATTT GTTCTTGAAC AGTCAAGAAA AACATTATTG AGGAAAATTA 1020

ATATCACAGC ATAACCCCAC CCTTTACATT TTGTGCAGTG ATTATTTTTT AAAGTCTTCT 1080

TTCATGTAAG TAGCAAACAG GGCTTTACTA TCTTTTCATC TCATTAATTC AATTAAAACC 1140

ATTACCTTAA AATTTTTTTC TTTCGAAGTG TGGTGTCTTT TATATTTGAA TTAGTAACTG 1200

TATGAAGTCA TAGATAATAG TACATGTCAC CTTAGGTAGT AGGAAGAATT ACAATTTCTT 1260

TAAATCATTT ATCTGGATTT TTATGTTTTA TTAGCATTTT CAAGAAGACG GATTATCTAG 1320

AGAATAATCA TATATATGCA TACGTAAAAA TGGACCACAG TGACTTATTT GTAGTTGTTA 1380

GTTGCCCTGC TACCTAGTTT GTTAGTGCAT TTGAGCACAC ATTTTAATTT TCCTCTAATT 1440

AAAATGTGCA GTATTTTCAG TGTCAAATAT ATTTAACTAT TTAGAGAATG ATTTCCACCT 1500

TTATGTTTTA ATATCCTAGG CATCTGCTGT AATAATATTT TAGAAAATGT TTGGAATTTA 1560

AGAAATAACT TGTGTTACTA ATTTGTATAA CCCATATCTG TGCAATGGAA TATAAATATC 1620

ACAAAGTTGT TTAACTAGAC TGCGTGTTGT TTTTCCCGTA TAATAAAACC AAAGAATAGT 1680

TTGGTTCTTC AAATCTTAAG AGAATCCACA TAAAAGAAGA AACTATTTTT TAAAAATTCA 1740

CTTCTATATA TACAATGAGT AAAATCACAG ATTTTTTCTT TAAATAAAAA TAAGTCATTT 1800

TAATAACTAA ACCAGATTCT TTGTGATACT ATTAANGTAA CATTTAGCCC CAAAAAAAAA 1860

AAA 1863

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

GGCACAGCGG CACGAGGTGA GCTGAGCCGG TGGGTGAGCG GCGGCCACGG CATCCTGTGC 60

TGTGGGGGCT ACGAGGAAAG ATCTAATTAT CATGGACCTG CGACAGTTTC TTATGTGCCT 120

GTCCCTGTGC ACAGCCTTTG CCTTGAGCAA ACCCACAGAA AAGAAGGACC GTGTACATCA 180
TGAGCCTCAG CTCAGTGACA AGGTTCACAA TGATGCTCAG AGTTTTGATT ATGACCATGA 240
TGCCTTCTTG GGTGCTGAAG AAGCAAAGAC CTTTGATCAG CTGACACCAG AAGAGAGCAA 300
GGAAAGGCTT GGAAAGATTG TAAGTAAAAT AGATGGCGAC AAGGACGGGT TTGTCACTGT 360
GGATGAGCTC AAAGACTGGA TTAAATTTGC ACAAAAGCGC TGGATTTACG AGGATGTAGA 420
GCGACAGTGG AAGGGGCATG ACCTCAATGA GGACGGCCTC GTTTCCTGGG AGGAGTATAA 480
AAATGCCACC TACGGCTACG TTTTAGATGA TCCAGATCCT GATGATGGAT TTAACTATAA 540
ACAGATGATG GTTAGAGATG AGCGGAGGTT TAAAATGGCA GACAAGGATG GAGACCTCAT 600
TGCCACCAAG GAGGAGTTCA CAGCTTTCCT GCACCCTGAG GAGTATGACT ACATGAAAGA 660
TATAGTAGTA CAGGAAACAA TGGAAGATAT AGATAAGAAT GCTGATGGTT TCATTGATCT 720
AGAAGAGTAT ATTGGTGACA TGTACAGCCA TGATGGGAAT ACTGATGAGC CAGAATGGGT 780
AAAGACAGAG CGAGAGCAGT TTGTTGAGTT TCGGGATAAG AACCGTGATG GGAAGATGGA 840
CAAGGAAGAG ACCAAAGACT GGATCCTTCC CTCAGACTAT GATCATGCAG AGGCAGAAGC 900
CAGGCACCTG GTCTATGAAT CAGACCAAAA CAAGGATGGC AAGCTTACCA AGGAGGAGAT 960
CGTTGACAAG TATGACTTAT TTGTTGGCAG CCAGGCCACA GATTTTGGGG AGGCCTTAGT 1020
ACGGCATGAT GAGTTCTGAG CTRCGGAGGA ACCCTCATTT CCTCAAAAGT AATTTATTTT 1080
TACAGCTTCT GGTTTCACAT GAAATTGTTT GCGCTACTGA GACTGTTACT ACAAACTTTT 1140
TAAGACATGA AAAGGCGTAA TGAAAACCAT CCCGTCCCCA TTCCTCCTCC TCTCTGAGGG 1200
ACTGGAGGGA AGCCGTGCTT CTGAGGAACA ACTCTAATTA GTACACTTGT GTTTGTAGAT 1260
TTACACTTTG TATTATGTAT TAACATGGCG TGTTTATTTT TGTATTTTTC TCTGGTTGGG 1320
AGTATGATAT GAAGGATCAA GATCCTCAAC TCACACATGT AGACAAACAT TAGCTCTTTA 1380
CTCTTTCTCA ACCCCTTTTA TGATTTTAAT AATTCTCACT TAACTAATTT TGTAAGCCTG 1440
AGATCAATAA GAAATGTTCA GGAGAGAGGA AAGAAAAAAA ATATATGCTC CACAATTTAT 1500
ATTTAGAGAG AGAACACTTA GTCTTGCCTG TCAAAAAGTC CAACATTTCA TAGGTAGTAG 1560
GGGCCACATA TTACATTCAG TTGCTATAGG TCCAGCAACT GAACCTGCCA TTACCTGGGC 1620
AAGGAAAGAT CCCTTTGCTC TAGGAAAGCT TGGCCCAAAT TGATTTTCTT CTTTTTCCCC 1680
CTGTAGGACT GACTGTTGGC TAATTTTGTC AAGCACAGCT GTGGTGGGAA GAGTTAGGGC 1740
CAGTGTCTTG AAAATCAATC AAGTAGTGAA TGTGATCTCT TTGCAGAGCT ATAGATAGAA 1800
ACAGCTGGAA AACTAAAGGA AAAATACAAG TGTTTTCGGG GCATACATTT TTTTTCTGGG 1860
TGTGCATCTG TTGAAATGCT CAAGACTTAA TTATTTGCCT TTTGAAATCA CTGTAAATGC 1920

CCCCATCCGG TTCCTCTTCT TCCCAGGTGT GCCAAGGAAT TAATCTTGGT TTCACTACAA 1980

TTAAAATTCA CTCCTTTCCA ATCATGTCAT TGAAAGTGCC TTTAACGAAA GAAATGGTCA 2040

CTGAATGGGA ATTCTCTTAA GAAACCCTGA GATTAAAAAA AGACTATTTG GATAACTTAT 2100

AGGAAAGCCT AGAACCTCCC AGTAGAGTGG GGATTTTTTT CTTCTTCCCT TTCTCTTTTG 2160

GACAATAGTT AAATTAGCAG TATTAGTTAT GAGTTTGGTT GCAGTGTTCT TATCTTGTGG 2220

GCTGATTTCC AAAAACCACA TGCTGCTGAA TTTACCAGGG ATCCTCATAC CTCACAATGC 2280

AAACCACTTA CTACCAGGCC TTTTTCTGTG TCCACTGGAG AGCTTGAGCT CACACTCAAA 2340

GATCAGAGGA CCTACAGAGA GGGCTCTTTG GTTTGAGGAC CATGGCTTAC CTTTCCTGCC 2400

TTTGACCCAT CACACCCCAT TTCCTCCTCT TTCCCTCTCC CCGCTGCCAA TTCCTGCAGC 2460

CCGGGGGAAC CACTAGTT 2478

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2058 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TCGGCCTTGC TTTTGTGGYC TTCCTCTGTG GCCAGAGCGT TTTCATCACC AAGCCTCCTG 60

ATGGCAGTNC CTTCACCGAT ATGTTCAAGA TACTGACGTA TTCCTGCTGT TCCCAGAAGC 120

GAAGTGGAGA GCGCCAGAGT AATGGTGAAG GCATTGGAGT NTTTCAGCAA TCTTCTAAAC 180

AAAGTCTGTT TGATTCATGT AAGATGTCTC ATGGTGGGCC ATTTACAGAA GAGAAAGTGG 240

AAGATGTGAA AGCTCTGGTC AAGATTGTCC CTGTTTTCTT GGCTTTGATA CCTTACTGGA 300

CAGTGTATTT CCAAATGCAG ACAACATATG TTTTACAGAG TCTTCATTTG AGGATTCCAG 360

AAATTTCAAA TATTACAACC ACTCCTCACA CGCTCCCTGC AGCCTGGCTG ACCATGTTTG 420

ATGCTGTGCT CATCCTCCTG CTCATCCCTC TGAAGGACAA ACTGGTCGAT CCCATTTTGA 480

GAAGACATGG CCTGCTCCCA TCCTCCCTGA AGAGGATCGC CGTGGGCATG TTCTTTGTCA 540

TGTGCTCRGC CTTTGCTGCA GGAATTTTGG AGAGTAAAAG GCTGAACCTT GTTAAAGAGA 600

AAACCATTAA TCAGACCATC GGCAACGTCG TCTACCATGC TGCCGATCTG TCGCTGTGGT 660

GGCAGGTGCC GCAGTACTTG CTGATTGGGA TCAGCGAGAT CTTTGCAAGT ATCGCAGGCC 720

TGGAATTTGC ATACTCAGCT GCCCCCAAGT CCATGCAGAG TGCCATAATG GGCTTGTTCT 780

TTTTCTTCTC TGGCGTCGGG TCGTTCGTGG GTTCTGGACT GCTGGCACTG GTGTCTATCA 840

AAGCCATCGG ATGGATGAGC AGTCACACAG ACTTTGGTAA TATTAACGGC TGCTATTTGA 900

ACTATTACTT TTTCCTTCTG GCTGCTATTC AAGGAGCTAC CCTCCTGCTT TTCCTCATTA 960

TTTCTGTGAA ATATGACCAT CATCGAGACC ATCAGCGATC AAGAGCCAAT GGCGTGCCCA 1020

CCAGCAGGAG GGCCTGACCT TCCTGAGGCC ATGTGCGGTT TCTGAGGCTG ACATGTCAGT 1080

AACTGACTGG GGTGCACTGA GAACAGGCAA GACTTTAAAT TCCCATAAAA TGTCTGACTT 1140

CACTGAAACT TGCATGTTGC CTGGATTGAT TTCTTCTTTC CCTCTATCCA AAGGAGCTTG 1200

GTAAGTGCCT TACTGCAGCG TGTCTCCTGG CACGCTGGGC CCTCCGGGAG GAGAGCTGCA 1260

GATTTCGAGT ATGTCGCTTG TCATTCAAGG TCTCTGTGAA TCCTCTAGCT GGGTTCCCTT 1320

TTTTACAGAA ACTCACAAAT GGAGATTGCA AAGTCTTGGG GAACTCCACG TGTTAGTTGG 1380

CATCCCAGTT TCTTAAACAA ATAGTATCAC CTGCTTCCCA TAGCCATATC TCACTGTAAA 1440

AAAAAAAATT AATAAACTGT TACTTATATT TAAGAAAGTG AGGATTTTTT TTTTTTAAAG 1500

ATAAAAGCAT GGTCAGATGC TGCAAGGATT TTACATAAAT GCCATATTTA TGGTTTCCTT 1560

CCTGAGAACA ATCTTGCTCT TGCCATGTTC TTTGATTTAG GCTGGTAGTA AACACATTTC 1620

ATCTGCTGCT TCAAAAAGTA CTTACTTTTT AAACCATCAA CATTACTTTT CTTTCTTAAG 1680

GCAAGGCATG CATAAGAGTC ATTTGAGACC ATGTGTCCCA TCTCAAGCCA CAGAGCAACT 1740

CACGGGGTAC TTCACACCTT ACCTAGTCAG AGTGCTTATA TATAGCTTTA TTTTGGTACG 1800

ATTGAGACTA AAGACTGATC ATGGTTGTAT GTAAGGAAAA CATTCTTTTG AACAGAAATA 1860

GTGTAATTAA AAATAATTGA AAGTGTTAAA TGTGAACTTG AGCTGTTTGA CCAGTCACAT 1920

TTTTGTATTG TTACTGTACG TGTATCTGGG GCTTCTCCGT TTGTTAATAC TTTTTCTGTA 1980

TTTGTTGCTG TATTTTTGGC ATAACTTTAT TATAAAAAGC ATCTCAAATG CGAAAWAAAA 2040

AAAAAAAAAA AAAAAAAC 2058

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

GGCACAGGAG CGACCCGGGA GAAGGAGGGC CAMGAKGCGG AAGCGGAGGA GTCTCCAGGA 60

GACCCGGGGA CAGCATCGCC CAGGCCCCTG TTTGCAGGCC TTTCAGATAT ATCCATCTCA 120

CAAGACATCC CCGTAGAAGG AGAAATCACC ATTCCTATGA GATCTCGCAT CCGGGAGTTT 180

GACAGCTCCA CATTAAATGA ATCTGTTCGC AATACCATCA TGCGTGATCT AAAAGCTGTT 240

GGGAAAAAAT TCATGCATGT TTTGTACCCA AGGAAAAGTA ATACTCTTTT GAGAGATTGG 300

GATTTGTGGG GCCCTTTGAT CCTTTGTGTG ACACTCGCAT TAATGCTGCA AAGAGACTCT 360

GCAGATAGTG AAAAAGATGG AGGGCCCCAA TTTGCAGAGG TGTTTGTCAT TGTCTGGTTT 420

GGTGCAGTTA CCATCACCCT CAACTCAAAA CTTCTTGGAG GGAACATATC TTTTTTTCAG 480

AGCCTCTGTG TGCTGGGTTA CTGTATACTT CCCTTGACAG TAGCAATGCT GATTTGCCGG 540

CTGGTACTTT TGGCTGATCC AGGACCTGTA AACTTCATGG TTCGGCTTTT TGTGGTGATT 600

GTGATGTTTG CCTGGTCTAT AGTTGCCTCC ACAGCTTTCC TTGCTGATAG CCAGCCTCCA 660

AACCGCAGAG CCCTAGCTGT TTATCCTGTT TTCCTGTTTT ACTTTGTCAT CAGTTGGATG 720

ATTCTCACCT TTACTCCTCA GTAAATCAGG AATGGGAAAT TAAAAACCAG TGAATTGAAA 780

GCACATCTGA AAGATGCAAT TCACCATGGA GCTTTGTCTC TGGCCCTTAT TTGTCTAATT 840

TTGGAGGTAT TTGATAACTG AGTAGGTGAG GAGATTAAAA GGGAGCCATA TAGCACTGTC 900

ACCCCTTATT TGAGGAACTG ATGTTTGAAA GGCTGTTCTT TTCTCTCTTA ATGTCATTTC 960

TTTAAAAATA CATGTGCATA CTACACACAG TATATAATGC CTCCTTAAGG CATGATGGAG 1020

TCACCGTGGT CCATTTGGGT GACAACCAGT GACTTGGGAA GCACATAGAT ACATCTTACA 1080

AGTTGAATAG AGTTGATAAC TATTTTCAGT TTTGAGAATA CCAGTTCAGG TGCAGCTCTT 1140

AAACACATTG CCTTATGACT ATTAGAATAT GCCTCTCTTT TCATAAATAA AAATACATGG 1200

TCTATATCCA TTTTCTTTTA TTTCTCTCTC TTAAGCTTAA AAAGGCAATG AGAGAGGTTA 1260

GGAGTGGGTT CATACACGGA GAATGAGAAA ACATGCATTA ACCAATATTC AGATTTTGAT 1320

CAGGGGAAAT TCTAYACTTG TTGCAAAAAA AAAAAAAAAA AAACTCGAGG GGGGCCCGGT 1380

ACCCAATCGC NGTATATGAT CGNAAACAAT C 1411

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2187 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

GCTTTGGCTT TTTTTGGCGG ACTGGGGCGC CCTCCGGAAG CGTTTCCAAC TTTCCAGAAG 60

TTTCTCGGGA CGGGCAGGAG GGGGTGGGGA CTGCCATATA TAGATCCCGG GAGCAGGGGA 120

GCGGGCTAAG AGTAGAATCG TGTCGCGCTC GAGAGCGAGA GTCACGTCCC GGCGCTAGCC 180

CAGCCCGACC CAGGCCCACC GTGGTGCACG CAAACCACTT CCTGGCCATG CGCTCCCTCC 240

TGCTTCTCAG CGCCTTCTGC CTCCTGGAGG CGGCCCTGGC CGCCGAGGTG AAGAAACCTG 300

CAGCCGCAGC AGCTCCTGGC ACTGCGGAGA AGTTGAGCCC CAAGGCGGCC ACGCTTGCCG 360

AGCGCAGCCG GCCTGGCCTT CAGCTTGTAC CAGGCCATGG CCAAGGACCA GGCAGTGGAG 420

AACATCCTGG TGTCACCCGT GGTGGTGGCC TCGTCGCTGG GGCTCGTGTC GCTGGGCGGC 480

AAGGCGACCA CGGCGTCGCA GGCCAAGGCA GTGCTGAGCG CCGAGCAGCT GCGCGACGAG 540

GAGGTGCACG CCGGCCTGGG CGAGCTGCTG CGCTCACTCA GCAACTCCAC GGCGCGCAAC 600

GTGACCTGGA AGCTGGGCAG CCGACTGTAC GGACCCAGCT CAGTGAGCTT CGCTGATGAC 660

TTCGTGCGCA GCAGCAAGCA GCACTACAAC TGCGAGCACT CCAAGATCAA CTTCCGCGAC 720

AAGCGCAGCG CGCTGCAGTC CATCAACGAG TGGGCCGCGC AGACCACCGA CGGCAAGCTG 780

CCCGAGGTCA CCAAGGACGT GGAGCGCACG GACGGCGCCC TGTTAGTCAA CGCCATGTTC 840

TTCAAGCCAC ACTGGGATGA GAAATTCCAC CACAAGATGG TGGACAACCG TGGCTTCATG 900

GTGACTCGGT CCTATACCGT GGGTGTCATG ATGATGCACC GGACAGGCCT CTACAACTAC 960

TACGACGACG AGAAGGAAAA GCTGCAAATC GTGGAGATGC CCCTGGCCCA CAAGCTCTCC 1020

AGCCTCATCA TCCTCATGCC CCATCACGTG GAGCCTCTCG AGCGCCTTGA AAAGCTGCTA 1080

ACCAAAGAGC AGCTGAAGAT CTGGATGGGG AAGATGCAGA AGAAGGCTGT TGCCATCTCC 1140

TTGCCCAAGG GTGTGGTGGA GGTGACCCAT GACCTGCAGA AACACCTGGC TGGGCTGGGC 1200

CTGACTGAGG CCATTGACAA GAACAAGGCC GACTTGTCAC GCATGTCAGG CAAGAAGGAC 1260

CTGTACCTGG CCAGCGTGTT CCACGCCACC GCCTTTGAGT TGGACACAGA TGGCAACCCT 1320

TTGACCAGAA TTACGGGCGG AGGAGTGCGC ACCCAAGTGT TCTACGCCGA CCACCCCTTC 1380

ATTTCCTAGT GCGGGACACC CAAAGCGGTC CCTGCTATTC ATTGGGCGCC TGGTCCGGCC 1440

TAAGGGTGAC AAGATGCGAG ACGAGTTATA GGCCTCAGGG TGCACACAGG ATGGCAGGAG 1500

GCATCCAAAG GCTCCTGAGA CACATGGGTG CTATTGGGGT TGGGGGGGAG GTGAGGTACC 1560

AGCCTTGGAT ACTCCATGGG GTGGGGTGGA AAAGCAGACC GGGGTTCCCG TGTGCCTGAG 1620

CGGACTTCCC AGCTAGAATT CACTCCACTT GGACATGGGC CCCAGATACC ATGATGCTGA 1680

GCCCGGAAAC TCCACATCCT GTGGGACCTG GGCCATAGTC ATTCTGCCTG CCCTGAAAGT 1740

CCCAGATCAA GCCTGCCTCA ATCAGTATTC ATATTTATAG CCAGGTACCT TCTCACCTGT 1800

GAGACCAAAT TGAGCTAGGG GGGTCAGCCA GCCCTCTTCT GACACTAAAA CACCTCAGCT 1860

GCCTCCCCAG CTCTATCCCA ACCTCTCCCA ACTATAAAAC TAGGTGCTGC AGCCCCTGGG 1920

ACCAGGCACC CCCAGAATGA CCTGGCCGCA GTGAGGCGGA TTGAGAAGGA GCTCCCAGGA 1980

GGGGCTTCTG GGCAGACTCT GGTCAAGAAG CATCGTGTCT GGCGTTGTGG GGATGAACTT 2040

TTTGTTTTGT TTCTTCCTTT TTTAGTTCTT CAAAGATAGG GAGGGAAGGG GGAACATGAG 2100

CCTTTGTTGC TATCAATCCA AGAACTTATT TGTACATTTT TTTTTTCAAT AAAACTTTTC 2160

CAATGACAAA AAAAAAAAAA AAAAAAA 2187

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 757 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

GACAGTACGG TCGGATTCCC GGGTCGACCC ACGCGTCCGC GGACGGTGAA GAAGGTGAAG 60

ATGGCGGTGG CCAGGGCCGG GGTCTTGGGA GTCCAGTGGC TGCAAAGGGC ATCCCGGAAC 120

GTGATGCCGC TGGGCGCACG GACAGCCTCC CACATGACCA AGGACATGTT CCCGGGGCCC 180

TATCCTAGGA CCCCAGAAGA ACGGGCCGCC GCCGCCAAGA AGTATAATAT GCGTGTGGAA 240

GACTACGAAC CTTACCCGGA TGATGGCATG GGGTATGGCG ACTACCCGAA GCTCCCTGAC 300

CGCTCACAGC ATGAGAGAGA TCCATGGTAT AGCTGGGACC AGCCGGGCCT GAGGTTGAAC 360

TGGGGTGAAC CGATGCACTG GCACCTAGAC ATGTACAACA GGAACCGTGT GGATACATCC 420

CCCACACCTG TTTCTTGGCA TGTCATGTGT ATGCAGCTCT TCGGTTTCCT GGCTTTCATG 480

ATATTCATGT GCTGGGTGGG GGACGTGTAC CCTGTCTACC AGCCTGTGGG ACCAAAGCAG 540

TATCCTTACA ATAATCTGTA CCTGGAACGA GGCGGTGATC CCTCCAAAGA ACCAGAGCGG 600

GTGGTTCACT ATGAGATCTG AGGAGGCTTC GTGGGCTTTT GGGTCCTCTA ACTAGGACTC 660

CCTCATTCCT AGAAATTTAA CCTTAATGAA ATCCCTAATA AAACTCAGTG CTGTGTTAAA 720

AAAAAAAAAA AAAAAAAAAA AAAAAGGGGG GCCCCNN 757

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

```
GGCACGAGCA CTCCTGCACT TCCCCACCCC CACGACCGAA CCTGGCTTCG CTAACGCCCT     60
CCCAGCTCCC TCGGGCCTGA CTTCCGGTTT CCTCGCGCGT CCCTGGCGCC GAGCCGCGGA    120
CAGCAGCCCC TTTTCCGGCT GAGAGCTCAT CCACACTTCC AATCACTTTC CGGAGTGCTT    180
CCCCTCCCTC CGGCCCGTGC TGGTCCCGAC GGCGGGCCTG GGTCTCGCGC GCGTATTGCT    240
GGGTAACGGG CCTTCTCYCG CGTCGGCCCG GCCCCTCCTG CCTCGGCTCG TCCCTCCTTC    300
CAGAACGTCC CGGGCTCCTG CCGAGTCAGA AGAAATGGGA CTCCCTCCGC GACGTGCCCG    360
GAGCAGCTCC CTTCGCTGTG GAAGCGGCGG TGTCTTCGAA GAAACCGGAA GCCCGTGGTG    420
ACCCCTGGCG ACCCGGTTTG TTTTCGGTCC GTTTCCAAAC ACTAAGGAAT CGAAACTCGG    480
CGGCCTTGGG GGCGGCCCTA CGTAGCCTGG CTTCTGGTTG TCATGGATGC ACTGGTAGAA    540
GATGATATCT GTATTCTGAA TCATGAAAAA GCCCATAAGA GAGATACAGT GACTCCAGTT    600
TCAATATATT CAGGAGATGA ATCTGTTGCT TCCCATTTTG CTCTTGTCAC TGCATATGAA    660
GACATCAAAA AACGACTTAA GGATTCAGAG AAAGAGAACT CTTTGTTAAA GAAGAGAATA    720
AGATTTTTGG AAGAAAAGCT AATAGCTCGA TTTGAAGAAG AAACAAGTTC CGTGGGACGA    780
GAACAAGTAA ATAAGGCCTA TCATGCATAT CGAGAGGTTT GCATTGATAG AGATAATTTG    840
AAGAGCAAAC TGGACAAAAT GAATAAAGAC AACTCTGAAT CTTTGAAAGT ATTGAATGAG    900
CAGCTACAAT CTAAAGAAGT AGAACTCCTC CAGCTGAGGA CAGAGGTGGA AACTCAGCAG    960
GTGATGAGGA ATTTAAATCC ACCTTCATCA AACTGGGAGG TGGAAAAGTT GAGCTGTGAC   1020
CTGAAGATCC ATGGTTTGGA ACAAGAGCTG GAACTGATGA GGAAAGAATG TAGCGATCTC   1080
AAAATAGAAC TACAGAAAGC CAAACAAACG GATCCATATC AGGAAGACAA TCTGAAGAGC   1140
AGAGATCTCC AAAAACTAAG CATTTCAAGT GATAATATGC AGCATGCATA CTGGGAACTG   1200
AAGAGAGAAA TGTCTAATTT ACATCTGGTG ACTCAAGTAC AAGCTGAACT ACTAAGAAAA   1260
CTGAAAACCT CAACTGCAAT CAAGAAAGCC TGTGCCCCTG TAGGATGCAG TGAAGACCTT   1320
GGAAGAGACA GCACAAAACT GCACTTGATG AATTTTACTG CAACATACAC AAGACATCCC   1380
CCTCTCTTAC CAAATGGCAA AGCTCTTTGT CATACCACAT CTTCCCCTTT ACCAGGAGAT   1440
GTAAAGGTTT TATCAGAGAA AGCAATCCTC CAATCATGGA CAGACAATGA GAGATCCATT   1500
CCTAATGATG GTACATGCTT TCAGGAACAC AGTTCTTATG GCAGAAATTC TCTGGAAGAC   1560
AATTCCTGGG TATTTCCAAG TCCTCCTAAA TCAAGTGAGA CAGCATTTGG GGAAACTAAA   1620
ACTAAAACTT TGCCTTTACC CAACCTTCCA CCACTGCATT ACTTGGATCA ACATAATCAG   1680
AACTGCCTTT ATAAGAATTA ATTTGGAAGA GATTCACGAT TTCACCATGA GGACACTTAT   1740
```

CTCTTTCAGT GGTCCTCCCA AGAAATTATT TAACAAACTG AANGGAGATT TTGATTAAAA 1800

TTTTGCAGAG GTCTTCAGTA TCTATATTTG AACACACTGT ACAATAGTAC AAAAACCAAC 1860

ATAGTTGGTT TTCTAGTATG AAAGAGCACC CTCTAGCTCC ATATTCTAAG AATCTGAAAT 1920

ATGCTACTAT ACTAATTAAT AAGTAAACTT AAGGTGTTTA AAAAACTCTG CCTTCTATAT 1980

TAATTGTAAA ATTTTGCCTC TCAGAAGAAT GGAATTGGAG ATTGTAGACG TGGTTTTACA 2040

AAATGTGAAA TGTCTAAATA TCTGTTCATA AAAATAAAAG GAAAACATGT TTCTTCAAAT 2100

TGCATAATGG AACAAATGGC AATGTGAGTA GGTTACATTT CTGTTGTTAT AATGCGTAAA 2160

GATATTGAAA ATATAATGAA ATAAAAGCAT CTTAGGTTAT ACCATCTTTA TATGCTATTG 2220

CGTTTCAATA TTTAAGATTT AAAGTGATTT TTTGGTCACA GTGTTTTGTT GATAAAATTT 2280

TTTTAGAATT GAAGTTTGAA TTCTAAGACT TGAAACAACC TGATCACTGA AGCCAACTTT 2340

GTCCCAGCAC ATTCCTTAAG TCCTAATTGG GGAAAAAAAA AAAAAAAAAC TCGA 2394

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 672 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NC: 96:

AGTGCTCTGT TGCCCAGGCT GGAGTGCGTT AGTGTAATGT CAGTCCACTG CAACCTCCAC 60

CCCCAGGTTC AAGCAATTCT CATGCCTCAG CCTCCCAAGT AGCTGAAATT ACTGGCATGC 120

ACCACCACAC CCAGCTGATG TTTATTTATT TATTTATATA TTTATTTATT TTAGGTGTTT 180

TTTTTTTTTT TTTTTGAGAC GGAGTCTTGC TCTGTTGCCC TGGGTGTGGT TACGTGGRAT 240

TACCATYCTG GGTGACTCAC TGAAATGTAC TCMCAGTGAG TCATGCCTTC MAATGACATC 300

TCAAGTTCTG CCTGCTTGGA GATACATCTG GGGATCTTAA GGGGTGAGGG ACTACTCAAC 360

AAGAAGGAAT TTAGCCTGTC TTTTTAAATA AACGGCATTT CTTTTTCCTA KAAAAATGGG 420

AAATTCTTCA ATTCTCTAAT ACAGGGACAC TGAGATAACA AAGAGGAAAG TGTCTGGTTG 480

GAGGTTGGGA RGCCACCCTG GGGTCTCTCC TACAAAAATG GAAAAGAAAA GAACGGTGAR 540

AAATCMAGCA AAGCACAARA AAKTTTCCCT TTGCTAAAAG GGAAAAGATG CCCCMCAATG 600

CCCATAAACA TGAACTGGGG ATAAGGAGGA RAATGTCTCT YCTTGGCACC CCCAAACAAA 660

CGTTAATTAC CC 672

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1419 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

```
TAAGAACAGA ACAGCAAGTA TGAACCACAT GGAACTTAAA ACATATGGGT GTGAAGTCCA     60
CTTATGTAGA CAAAACTTAT AATTTCCAAA CTGTTGTCTA GTATACAGTG ATCAGTTGCT    120
CTCTGTTCAA GTCATTCCAC ACATTTCCCT ATTTTAGGCT ATTATAATAT AGAAAGAAAA    180
TGGGAAGCAT TAGTTGGAGC TAGAAAATGA ACTGTATATT ATTGCTATAT TTGCTAATAC    240
CAACTATTTC AATAAGTGTT GTACCATATG TAGCATTAAA TATAAAATAC ATAAAAGAAT    300
GTACAGAAAA TAGCTTTTAT TGAGTAATAT TACATTTCAT TTATACTGTA GCAATATATT    360
TGTAGGTATA CTCTGTAAGG GCTTTAAATA AAAGAGGTCC ATTAATACTT CCTTATAAAA    420
ATTCTAGTCT GTTTCATTAC TGCCCAGATG TTTTAGAGAT AAATATTTAT GCAGAAGGTA    480
TTTTKGAAAG TCYCCYTTTG TCTGATAGAG TTTAACNAGA TATTTAAATT TAGTGCYCNA    540
GAAATCCCAC AAGTCACGGT CTAAACACAC TTAGAATACT ACAGCATAAA TCTGTTAGCA    600
TTANTTGCCA AATAAGACAG TTGGGATCCC AAACCCCAAG TCCTTGAGCA ATGTTTTTCC    660
TCAAAAAGCT GCTATNCCAA TGATATAGGA AAAWACATTG TGTTTTCCTA AACACACTTT    720
TCTTTTTAAA TGTGCTTCAT TGTTTGATTT GGTCCTGCCT AAATTTCACA AGCTAGGCCA    780
ATGAAGGCTG AATCAAAGAC ATTTCATCCA CCAATATCAT GTGTAGATAT TATGTATAGA    840
AAATAAAATA AATTATGGCT CTAACTTCTG TGTTGCTGTT TATCTTGTTA TTTTTCGGCG    900
TTATACTAAT GNGTTTATTG AGAGCATTTT ACCTTCCAGA CTTCTCATGG CTAACTTTTG    960
GTCTGWATTT TGSTCCTTAG ATGKGAATAT TTCTTATTAG TYTGCTYCCT GCWACGCAAT   1020
GACTGCATTT CTATCATTTC TCAGTTTGTT AGWATATGTG GATAGTATTC TACTGTATAA   1080
ATGATTGCAA AGTTTATCAA AAACAAATTA TTATATGTAG CTTTTCTACA GTGCTTTGCT   1140
AAACCATGTA GTACTAGTTA AGTSTTCCTT GAAAATAAAG ATACACTCTT ATAGGGGACA   1200
GTTCCTGTTC ACTCCCAGGA AACTTTTTTA AAAGATGACA CTGAATGTTT ATTGCACTTT   1260
AGTGCAGTGA AGTGGCAATA AAACCTAACA TGAATCAAGG TTGTTTATGG CAGATGCATG   1320
TGTTGCTTTA CAGAGTTTAG CAAAAGCTCT TAATTTTATG TCATACTGTA TTCTACTGAA   1380
TAATAAAGCT AACATTATTC AATAATAAAA TGGAAAAAA                          1419
```

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1487 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

```
GCGACCGCGC CCCTTTCAGC TAGCTCGCTC GCTCGCTCTG CTTCCCTGCT GCCGGCTGCG     60
CATGGCKWTG GCGTTGGCGG CGCTGGCGGC GGTCGAGCCG GCCTGCGCAG CCGGTACCAG    120
CAGTTGCAGA ATGAAGAAGA GTCTGGAGAA CCTGAACAGG CTGCAGGTGA TGCTCCTCCA    180
CCTTACAGCA GCATTTCTGC AGAGAGCGCA GTTTTCCACC TATTTCCCTG GATATTTTGA    240
TGGTCAGTAC TGGCTCTGGT GGGTGTTCCT TGTTTTAGGC TTTCTCCTGT TTCTCAGAGG    300
ATTTATCAAT TATGCAAAAG TTCGGAAGAT GCCAGAAACT TTCTCAAATC TCCCCAGGAC    360
CAGAGTTCTC TTTATTTATT AAAGATGTTT TCTGGCAAAG GCCTTCCTGC ATTTATGAAT    420
TCTCTCTCAA GAAGCAAGAG AACACCTGCA GGAAGTGAAT CAAGATGCAG AACACAGAGG    480
AATAATCACC TGCTTTAAAA AAATAAAGTA CTGTTGAAAA GATCATTTCT CTCTATTTGT    540
TCCTAGGTGT AAAATTTTAA TAGTTAATGC AGAATTCTGT AATCATTGAA TCATTAGTGG    600
TTAATGTTTG AAAAAGCTCT TGCAATCAAG TCTGTGATGT ATTAATAATG CCTTATATAT    660
TGTTTGTAGT CATTTTAAGT AGCATGAGCC ATGTCCCTGT AGTCGGTAGG GGGCAGTCTT    720
GCTTTATTCA TCCTCCATCT CAAAATGAAC TTGGAATTAA ATATTGTAAG ATATGTATAA    780
TGCTGGCCAT TTTAAAGGGG TTTTCTCAAA AGTTAAACTT TTGTTATGAC TGTGTTTTTG    840
CACATAATCC ATATTTGCTG TTCAAGTTAA TCTAGAAATT TATTCAATTC TGTATGAACA    900
CCTGGAAGCA AAATCATAGT GCAAAAATAC ATTTAAGGTG TGGTCAAAAA TAAGTCTTTA    960
ATTGGTAAAT AATAAGCATT AATTTTTTAT AGCCTGTATT CACAATTCTG CGGTACCTTA   1020
TTGTACCTAA GGGATTCTAA AGGTGTTGTC ACTGTATAAA ACAGAAAGCA CTAGGATACA   1080
AATGAAGCTT AATTACTAAA ATGTAATTCT TGACACTCTT TCTATAATTA GCGTTCTTCA   1140
CCCCCACCCC CACCCCCACC CCCCTTATTT TCCTTTTGTC TCCTGGTGAT TAGGCCAAAG   1200
TCTGGGAGTA AGGAGAGGAT TAGGTACTTA GGAGCAAAGA AAGAAGTAGC TTGGAACTTT   1260
TGAGATGATC CCTAACATAC TGTACTACTT GCTTTTACAA TGTGTTAGCA GAAACCAGTG   1320
GGTTATAATG TAGAATGATG TGCTTTCTGC CCAAGTGGTA ATTCATCTTG GTTTGCTATG   1380
```

TTAAAACTGT AAATACAACA GAACATTAAT AAATATCTCT TGTGTAGCAC CTTTAAAAAA 1440

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAN CCCGGGGGGG GGCCCCN 1487

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1653 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

GCGACCGCGC CCTTCAGCTA GCTCGCTCGC TCGCTCTGCT TCCCTGCTGC CGGCTGCGCA 60

TGGCTTNGGC GTTGGCGGCG CTGGCGGCGG CTCGAGCCGC CTGCGSAGCC GGTACCAGCA 120

GTTGCAGAAT GAAGAAGAGT CTGGAGAACC TGAACAGGCT GCAGGTGATG CTCCTCCACC 180

TTACAGCAGC ATTTCTGCAG AGAGCGCACA TNATTTTGAC TACAAGGATG AGTCTGGGTT 240

TCCAAAGCCC CCATCTTACA ATGTAGCTAC AACACTGCCC AGTTATGATG AAGCGGAGAG 300

GACCAAGGCT GAAGCTACTA TCCCTTTGGT TCCTGGGAGA GATGAGGATT TTGTGGGTCG 360

GGATGATTTT GATGATGCTG ACCAGCTGAG GATAGGAAAT GATGGGATTT TCATGTTAAC 420

TTTTTTCATG GCATTCCTCT TTAACTGGAT TGGGTTTTTC CTGTCTTTTT GCCTGACCAC 480

TTCAGCTGCA GGAAGGTATG GGGCCATTTC AGGATTTGGT CTCTCTCTAA TTAAATGGAT 540

CCTGATTGTC AGGTTTTCCA CCTATTTCCC TGCATTTATG AATTCTCTCT CAAGAAGCAA 600

GAGAACACCT GCAGGAAGTG AATCAAGATG CAGAACACAG AGGAATAATC ACCTGCTTTA 660

AAAAAATAAA GTACTGTTGA AAAGATCATT TCTCTCTATT TGTTCCTAGG TGTAAAATTT 720

TAATAGTTAA TGCAGAATTC TGTAATCATT GAATCATTAG TGGTTAATGT TTGAAAAAGC 780

TCTTGCAATC AAGTCTGTGA TGTATTAATA ATGCCTTATA TATTGTTTGT AGTCATTTTA 840

AGTAGCATGA GCCATGTCCC TGTAGTCGGT AGGGGGCAGT CTTGCTTTAT TCATCCTCCA 900

TCTCAAAATG AACTTGGAAT TAAATATTGT AAGATATGTA TAATGCTGGC CATTTTAAAG 960

GGGTTTTCTC AAAAGTTAAA CTTTTGTTAT GACTGTGTTT TTGCACATAA TCCATATTTG 1020

CTGTTCAAGT TAATCTAGAA ATTTATTCAA TTCTGTATGA ACACCTGGAA GCAAAATCAT 1080

AGTGCAAAAA TACATTTAAG GTGTGGTCAA AAATAAGTCT TTAATTGGTA AATAATAAGC 1140

ATTAATTTTT TATAGCCTGT ATTCACAATT CTGCGGTACC TTATTGTACC TAAGGGATTC 1200

TAAAGGTGTT GTCACTGTAT AAAACAGAAA GCACTAGGAT ACAAATGAAG CTTAATTACT 1260

AAAATGTAAT TCTTGACACT CTTTCTATAA TTAGCGTTCT TCACCCCCAC CCCCACCCCC 1320

ACCCCCCTTA TTTTCCTTTT GTCTCCTGGT GATTAGGCCA AAGTCTGGGA GTAAGGAGAG 1380

GATTAGGTAC TTAGGAGCAA AGAAAGAAGT AGCTTGGAAC TTTTGAGATG ATCCCTAACA 1440

TACTGTACTA CTTGCTTTTA CAATGTGTTA GCAGAAACCA GTGGGTTATA ATGTAGAATG 1500

ATGTGCTTTC TGCCCAAGTG GTAATTCATC TTGGTTTGCT ATGTTAAAAC TGTAAATACA 1560

ACAGAACATT AATAAATATC TCTTGTGTAG CACCTTTTAW AAAAAAAAAA AAAAAAAAAA 1620

AAAAAAAAAA AAAAANCCCG GGGGGGGGCC CCN 1653

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1145 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

TTTTTTTTTT TTTTTTTTTT TTGACTGAAC TAAGTGGCTT TTTTATTAGA GAAAGCCAGA 60

ATTACAAAAG ACTTCCCTTT TCTTGGGGTA TGGCTGTCTC AGCACAATAC TCAACATAAC 120

TGCAGAACTG ATGTGGCTCA GGCACCCTGG TTTTAATTCC TTGAGGATCT GGCAATTGGC 180

TTACGCAAAA GGTCACCATT TGAGGTCCTG CCTTACTAAT TATGTGCTGC CCAACAACTA 240

AATTTGTAAT TTGTTTTTCT CTAGTTTGAG CAGGGTCTGA ATTTTTTCAT TTATTTCCTT 300

TTTTGCCAGC AGACAGACTT GAGTCTGTAA AGACAAGCAA ATACACTGAC AGAAGTTTAC 360

CATAGTTTCT AAAATGTAAA AAAGAAAACC CCCAAAAGAC TCAAGAAAAT TAGACCACAA 420

ATTTTGCATT GTTCATTGTA GCACTATTGG TAATAAAATA ACAAATGTTT GTGCATTTTT 480

ATGTGAAGAT CCTTCTCGTA TTTCATTTGG AAAGATGAGC AAGAGGTCTG CTTCCTTCAT 540

TTTACTTCCC CTTCTGTTTT TGAAAGGCAG TTTCGCCAAG CTTAATGCAA GAATATCTGA 600

CTGTTTAGAA GAAAGATATT GCCACAATCT CTGGATGGTT TTCCAGGGTT GTGTTATTAC 660

TGAGCTTCAT CTTTCCAGAA TGAGCAAAAC ACTGTCCAGT CTTTGTTACG ATTTTGTAAT 720

AAATGTGTAC ATTTTTTTTA AATTTTTGGA CATCACATGA ATAAAGGTAT GTATGTACGA 780

ATGTGTATAT ATTATATATA TGACATCTAT TTTGGAAAAT GTTTGCCCTG CTGTACCTCA 840

TTTTTAGGAG GTGTGCATGG ATGCAATATA TGAAAATGGG ACATTCTGGA ACTGCTGGTC 900

AGGGGACTTT GTCGCCCTGT GCACTAAAAG GGCCAGATTT TCAGCAGCCA AGGACATCCA 960

TACCCAAGTG AATGTGATGG GACTTAAAAG AAGTGAACTG AGACAATTCA CTCTGGCTGT 1020

TTGAACAGCA GCGTTTCATA GGAAGAGAAA AAAAGATCAA TCTTGTATTT TCTGACCACA 1080

TAAAGGCTTC TTCTCTTTGT AATAAAGTAG AAAAGCTCTC CTCAAAAAAA AAAAAAAAAA 1140

AAAAA 1145

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 734 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

TACCCGGCGG ATTCCAGGAA GGTAAATTTA GTCCTATAAT TTTCAGCTTA ATTATAAACA 60

AAGGAACAAA TAAGTGGAAG GGCAGCTATT ACCATTCGCT TAGTCAAAAC ATTCGGTTAC 120

TGCCCTTTAA TACACTCCTA TCATCAGCAC TTCCACCATG TATTACAAGT CTTGACCCAT 180

CCCTGTCGTA ACTCCAGTAA AAGTTACTGT TACTAGAAAA TTTTTATCAA TTAACTGACA 240

AATAGTTTCT TTTTAAAGTA GTTTCTTCCA TCTTTATTCT GACTAGCTTC CAAAATGTGT 300

TCCCTTTTTG AATCGAGGTT TTTTTGTTTT GTTTTGTTTT CTGAAAAAAT CATACAACTT 360

TGTGCTTCTA TTGCTTTTTT GTGTTTTGTT AAGCATGTCC CTTGGCCCAA ATGGAAGAGG 420

AAATGTTTAA TTAATGCTTT TTAGTTTAAA TAAATTGAAT CATTTATAAT AATCAGTGTT 480

AACAATTTAG TGACCCTTGG TAGGTTAAAG GTTGCATTAT TTATACTTGA GATTTTTTTC 540

CCCTAACTAT TCTGTTTTTT GTACTTTAAA ACTATGGGGG AAATATCACT GGTCTGTCAA 600

GAAACAGCAG TAATTATTAC TGAGTTAAAT TGAAAAGTCC AGTGGACCAG GCATTTCTTA 660

TATAAATAAA ATTGGTGGTA CTAATGTGAA AAAAAAAAAA AAAAAAAACT CGAGGGGGGC 720

CCGGTACCCT ATTA 734

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 713 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

CCGCGGGAAC GCTGTCCTGG CTGCCGNCAC CCGAACAGCC TGTCCTGGTG CCCCGGCTCC 60

CTGCCCCGCG CCCAGTCATG ACCCTGCGCC CCTCACTCCT CCCGCTCCAT CTGCTGCTGC 120

TGCTGCTGCT CAGTGCGGCG GTGTGCCGGG CTGAGGCTGG GCTCGAAACC GAAAGTCCCG 180

TCCGGACCCT CCAAGTGGAG ACCCTGGTGG AGCCCCCAGA ACCATGTGCC GAGCCCGCTG 240

CTTTTGGAGA CACGCTTCAC ATACACTACA CGGGAAGCTT GGTAGATGGA CGTATTATTG 300

ACACCTCCCT GACCAGAGAC CCTCTGGTTA TAGAACTTGG CCAAAAGCAG GTGATTCCAG 360

GTCTGGAGCA GAGTCTTCTC GACATGTGTG TGGGAGAGAA GCGAAGGGCA ATCATTCCTT 420

CTCACTTGGC CTATGGAAAA CGGGGATTTC CACCATCTGT CCCAGCGGAT GCAGTGGTGC 480

AGTATGACGT GGAGCTGATT GCACTAATCC GAGCCAACTA CTGGCTAAAG CTGGTGAAGG 540

GCATTTTGCC TCTGGTAGGG ATGGCCATGG TGCCACCCTC CTGGGCCTCA TTGGGTATCA 600

CCTATACAGA AAGGCCAATA GACCCAAAGT CTCCAAAAAG AAGCTCAAGG AAGAGAAACG 660

AAACAAGAGC AAAAAGAAAT AATAAATAAT AAATTTTAAA AAACTTAAAA AAA 713

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

CCGATGTGGA CATCATCCTG TCTATCCCCA TGTTCCTGCG CCTGTACCTG ATCGCCCGAG 60

TCATGCTGCT GCACAGAAGC TCTTCACCGA TGCCTCGTCC CGCAGCATCG GGGCCCTCAA 120

CAAGATCAAC TTCAACACCC GCTTTGTCAT GAAGACGCTC ATGACCATCT GCCCTGGCAC 180

TGTGCTGCTC GTGTTCAGCA TCTCTCTGTG GATCATTGCT GCCTGGACCG TCCGTGTCTG 240

TGAAAGTCCT GAATCACCAG CCCAGCCTTC TGGCTCATCA CTTCCTGCTT GGTACCATGA 300

CCAGCAGGAC GTAACTAGTA ACTTTCTGGG TGCCATGTGG CTCATCTCCA TCACATTCCT 360

TTCCATTGGT TATGGGGACA TGGTGCCCCA CACATACTGT GGGAAAGGTG TCTGTCTCCT 420

CACTGGCATC ATGGGTGCAG GCTGCACTGC CCTTGTGGTG GCCGTGGTGG CCCGAAAGCT 480

GGAACTCACC AAAGCGGAGA AGCACGTTCA TAANTTCATG ATGGACACTC AGCTCACCAA 540

GCGGATCAAG AATGYTGCAG CCAATGTCCT TSGGGAAACA TGGTTAATCT ATAAACACAC 600

AAAGYTGYTA AAGAAGATTG ACCATGCCAA AGTGAGGAAC ACCAGAGGAA GTTCYTCCAA 660

GTATCCACCA GTTGAGGAGC GTCAAGATGG AACAGAGGAA GCTGAGTGAC CAAGCCAACA 720

NTCTGGTGGA CCTTTCCAAG ATGCAGAATG TCMTGTATGA CTTAATCACA GAACTCAATG 780

ACCGGAGCGA AGACCTGGAG AAGCAGATTG GCAGCCTGGA GTCGAAGCTG GAGCATCTCA 840

CCGCCAGCTT CAACTCCCTG CCGCTGCTCA TCGCCGACAC CCTGCGCCAG CAGCAGCAGC 900

AGCTCCTGTC TGCCATCATC GAGGCCCGGG GTGTCAGCGT GGCAGTGGGC ACCACCCACA 960

CCCCAATCTC CGATAGCCCC ATTGGGGTCA GCTCCACCTC CTTCCCGACC CCGTACACAA 1020

GTTCAAGCAG TTGCTAAATA AATCTCCCCA CTCCAGAAGC ATTAAAAAAA AAAAAAAAAA 1080

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 489 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

GGCACGAGAG GCTTTGAAGC ATTTTTGTCT GTGCTCCCTG ATCTTCAGGT CACCACCATG 60

AAGTTCTTAG CAGTCCTGGT ACTCTTGGGA GTTTCCATCT TTCTGGTCTC TGCCCAGAAT 120

CCGACAACAG CTGCTCCAGC TGACACGTAT CCAGCTACTG GTCCTGCTGA TGATGAAGCC 180

CCTGATGCTG AAACCACTGC TGCTGCAACC ACTGCGACCA CTGCTGCTCC TACCACTGCA 240

ACCACCGCTG CTTCTACCAC TGCTCGTAAA GACATTCCAG TTTTACCCAA ATGGGTTGGG 300

GATCTCCCGA ATGGTAGAGT GTGTCCCTGA GATGGAATCA GCTTGAGTCT TCTGCAATTG 360

GTCACAACTA TTCATGCTTC CTGTGATTTC ATCCAACTAC TTACCTTGCC TACGATATCC 420

CCTTTATCTC TAATCAGTTT ATTTTCTTTC AAATAAAAAA TAACTATGAG CAACAAAAAA 480

AAAAAAAAA 489

(2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 640 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

GCGGTCGCCG CTGTTGTTGT GGTCCCCATG GAGCTGCCGT AGCGGACCCA GCACAGCCAG 60

GAGCGTCCGG GATGAGCTCA GCCGCGGCCG ACCACTGGGC GTGGTTGCTG GTGCTCAGCT 120

TCGTGTTTGG ATGCAATGTT CTTAGGATCC TCCTCCCGTC CTTCTCATCC TTCATGTCCA 180

GGGTGCTGCA GAAGGACGCG GAGCAGGAGT CACAGATGAG AGCGGAGATC CAGGACATGA 240

AGCAGGAGCT CTCCACAGTC AACATGATGG ACGAGTTTGC CAGATATGCC AGGCTGGAAA 300

GAAAGATCAA CAAGATGACG GATAAGCTCA AAACCCATGT GAAAGCTCGG ACAGCTCAAT 360

TAGCCAAGAT AAAATGGGTG ATAAGTGTCG CTTTCTACGT ATTGCAGGCT GCCCTGATGA 420

TCTCACTCAT TTGGAAGTAT TATTCTGTCC CTGTGGCTGT CGTGCCGAGT AAATGGATAA 480

CCCTYTAGAC CGCCTGGTAG CCTTTCCYAY TAGAGTAGCA GGTGGTGTTG GAATTACTGT 540

TGGATTTART CTGTACAAAT TGTCCTATTG TGCTTCACCG TYCASTGAAC AGGAGGTGGT 600

ACAGCCGGAG TTAAAAACGG TTTCCNTTCC AGTTTAAAAT 640

(2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1529 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

GGGCACNAGA TGGAGCTGCC GTAGCGGACC CAGCACAGCC AGGAGCGTCC GGGATGAGCT 60

CAGCCGCGGC CGACCACTGG GCGTGGTTGC TGGTGCTCAG CTTCGTGTTT GGATGCAATG 120

TTCTTAGGAT CCTCCTCCCG TCCTTCTCAT CCTTCATGTC CAGGGTGCTG CAGAAGGACG 180

CGGACAGGAG TCACAGATGA GAGCGGAGAT CCAGGACATG AAGCAGGAGC TCTCCACAGT 240

CAACATGATG GACGAGTTTG CCAGATATGC CAGGCTGGAA AGAAAGATCA ACAAGATGAC 300

GGATAAGCTC AAAACCCATG TGAAAGCTCG GACAGCTCAA TTAGCCAAGA TAAAATGGGT 360

GATAAGTGTC GCTTTCTACG TATTGCAGGC TGCCCTGATG ATCTCACTCA TTTGGAAGTA 420

TTATTCTGTC CCTGTGGCTG TCGTGCCGAG TAAATGGATA ACCCCTCTAG ACCGCCTGGT 480

AGCCTTTCCT ACTAGAGTAG CAGGTGGTGT TGGAATTACC TGTTGGATTT TAGTCTGTAA 540

CAAAGTTGTC GCTATTGTGC TTCATCCGTT CAGCTGAACA GGAGGATGGA TACAGCCGCG 600

AGTAAAAAAA CGGATTTCCT CTTCCTAGCT TAAAATCTGA TTTACACTGT TTTGTTTTTT 660

AAGAAACAAA AGTGCATAGT TTAGATTTTT TTTTTGTTGA ATATGTTTGT TCTTGGACTT 720

TATGAGATAG TCTTATAAGA ATCACGATTT TCTACACCTG TCATTGAGCC AAGAAAGTCC 780

AGTTTATGAC ACGTATGTAC TAGTGAACAC CGTCCTCGAT CTGTACGAAA TGTGAAATGT 840

TTAGGGACAT CTCCATGCTG TCACTTGTGA TTTGCCCTCT TATGTATTTT GGTCATATTG 900

CCAACTGGAA AGTCAAAATT TTCTAACAAC TTTAAGTAAG TTCTTTGAAG ACTTAGTGCT 960

GTTTTTAATC CAGTTTAGAA AGTAACTTAA TTTTAATACC RCTACTAAAA ATTCGAAAAT 1020

TTCTTCTTTA ATCACATTCA ATATGGTTAA AAGAACAACA CTAATTGACA TTGCGTGGGC 1080

TTTTTCTCCC TTTGTTTAAA ATGTCATTTG TTGAGCAAGA GTTGTATAGT ATTATCTACT 1140

TACTTGAGGC TGTTAATTTT TCATTACAGT GTTTTGTAAA TGTATCCACG AGACCATGAT 1200

GCATTGTTTT GTGCTCAACT TGTGTTTTGT ATTTAAAGCA TTTTGAATGA AGTGTATTTT 1260

ATAAGCATTT AATATTTATG CTCTTTAGAA TGGAACACAG AAAACAAACC TTATAAGTCC 1320

TGATTAATCT GAACCAATAA CCTGTGTGGC CTACAAAGTA TAATTCTATT AAATGTTCCT 1380

TAAAACACTT TTTTCTAATT AAAATCTTTG CAAATGCTTG TGTAACTTCC TGCCTTACAG 1440

CTACTTGTTT GCTGTGAGCC ACCCGCAACT GACAAGTGGC TGTTAACTGA GTCACCATAT 1500

CCCAGTAAAG CTGAATTTTC TCACTAAAA 1529

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

ATGAAGGGTC GTTGGTGGGA AAGATGGCGG CGACTCTGGG ACCCCTTGGT CGTGGCAGCA 60

GTGGCGRCGA TGTTTGTCGG CTCGGGATGG GTCCAGGATG TTACTCCTTC TTCTTTTGTT 120

GGGGTCTGGG CAGGGGCCAC AGCAAGTCGG GGCGGGTCAA ACGTTCGAGT ACTTGAAACG 180

GGAGCACTCG CTGTCGAAGC CCTACCAGGG TGTGGGCACA GGCAGTTCCT CACTGTGGAA 240

TCTGATGGGC AATGCCATGG TGATGACCCA GTATATCCGC CTTACCCCAG ATATGCAAAG 300

TAAACAGGGT GCCTTGTGGA ACCGGGTGCC ATGTTTCCTG AGAGACTGGG AGTTGCAGGT 360

GCACTTCAAA ATCCATGGAC AAGGAAAGAA GAATCTGCAT GGGGATGGCT TGGCAATCTG 420

GTACACAAAG GRWTCGGATG CAGCCAGGGC CTGTNTTTGG GAAACATGGA CAAATTTGTG 480

GGGCTGGGAG TATTTGTAGA CACCTACCCC AATGAGGAGA AGCAGCAAGA GCGGGTATTC 540

CCCTRCMTCT CAGCCATGGT GAACAACGGC TCCCTCAGCT ATGATCATGA GCGGGATGGG 600

CGGCCTACAG AGCTGGGAGG CTGCASAGCC ATTGTCCGCA ATCTTCATTA CGACACCTTC 660

CTGGTGATTC GCTACGTCAA GAGGCATTTR ACGATAATGA TGGATATTGA TGGCAAGCAT 720

GAGTGGAGGG ACTGCATTGA AGTGCCCGGA GTCCGCCTGC CCCGCGGCTA CTACTTCGGC 780

ACCTCCTCCA TCACTGGGGA TCTCTCAGAT AATCATGATG TCATTTCCTT GAAGTTGTTT 840
GAACTGACAG TGGAGAGAAC CCCAGAAGAG GAAAAGCTCC ATCGAGATGT GTTCTTGCCC 900
TCAGTGGACA ATATGAAGCT GCCTGAGATG ACAGCTCCAC TGCCGCCCCT GAGTGGCCTG 960
GCCCTCTTCC TCATCGTCTT TTTCTCCCTG GGTGTTTTCT GTATTTGCCA TAGTCATTGG 1020
TATCATACTC TACAACAAAT GGCAGGAACA GAGCCGAAAG CGCTTCTACT GAGCCCTCCT 1080
GCTGCCACCA CTTTTGTGAC TGTCACCCAT GAGGTATGGA AGGAGCAGGC ACTGGCCTGA 1140
GCATGCAGCC TGGAGAGTGT TCTTGTCTCT AGCAGCTGGT TGGGGACTAT ATTCTGTCAC 1200
TGGAGTTTTG AATGCAGGGA CCCCGCATTC CCATGGTTGT GCATGGGGAC ATCTAACTCT 1260
GGTCTGGGAA GCCACCCACC CCAGGGCAAT GCTGCTGTGA TGTGCCTTTC CCTGCAGTCC 1320
TTCCATGTGG GAGCAGAGGT GTGAAGAGAA TTTACGTGGT TGTGATGCCA AAATCACAGA 1380
ACAGAATTTC ATAGCCCAGG CTGCCGTGTT GTTTGACTCA GAAGGCCCTT CTACTTCAGT 1440
TTTGAATCCA CAAAGAATTA AAAACTGGTA ACACCACAGG CTTTCTGACC ATCCATTCGT 1500
TGGGTTTTGC ATTTGACCCA ACCCTCTGCC TACCTGAGGA GCTTTCTTTG GAAACCAGGA 1560
TGGAAACTTC TTCCCTGCCT TACCTTCCTT TCACTCCATT CATTGTCCTC TCTGTGTGCA 1620
ACCTGAGCTG GGAAAGGCAT TTGGATGCCT CTCTGTTGGG GCCTGGGGCT GCAGAACACA 1680
CCTGCGTTTC ACTGGCCTTC ATTAGGTGGC CCTAGGGAGA TGGCTTTCTG CTTTGGATCA 1740
CTGTTCCCTA GCATGGGTCT TGGGTCTATT GGCATGTCCA TGGCCTTCCC AATCAAGTCT 1800
CTTCAGGCCC TCAGTGAAGT TTGGCTAAAG GTTGGTGTAA AAATCAAGAG AAGCCTGGAA 1860
GACATCATGG ATGCCATGGA TTAGCTGTGC AACTGACCAG CTCCAGGTTT GATCAAACCA 1920
AAAGCAACAT TTGTCATGTG GTCTGACCAT GTGGAGATGT TTCTGGACTT GCTAGAGCCT 1980
GCTTAGCTGC ATGTTTTGTA GTTACGATTT TTGGAATCCC ACTTTGAGTG CTGAAAGTGT 2040
AAGGAAGCTT TCTTCTTACA CCTTGGGCTT GGATATTGCC CAGAGAAGAA ATTTGGCTTT 2100
TTTTTTNCTT AATGGACAAG AGACAGTTGC TGTTCTCATG TTCCAAGTCT GAGAGCAACA 2160
GACCCTCATC ATCTGTGCCT GGAAGAGTTC ACTGTCATTG AGCAGCACAG CCTGAGTGCT 2220
GGCCTCTGTC AACCCTTATT CCACTGCCTT ATTTGACAAG GGGTTACATG CTGCTCACCT 2280
TACTGCCCTG GGATTAAATC AGTTACAGGC CAGAGTCTCC TTGGAGGGCC TGGAACTCTG 2340
AGTCCTCCTA TGAACCTCTG TAGCCTAAAT GAAATTCTTA AAATCACCGA TGGAACCAAA 2400
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAN 2435

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 805 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

ATGAAACTTA AGAATTGAAT TGGAAAGACT TCTCAAAGAG AATTGTATGT AACGATGTTG 60

TATTGATTTT TAAGAAAGTA ATTTAATTTG TAAAACTTCT GCTCGTTTAC ACTGCACATT 120

GAATACAGGT AACTAATTGG AAGGAGAGGG GAGGTCACTC TTTTGATGGT GGCCCTGAAC 180

CTCATTCTGG TTCCCTGCTG CGCTGCTTGG TGTGACCCAC GGAGGATCCA CTCCCAGGAT 240

GACGTGCTCC GTAGCTCTGC TGCTGATACT GGGTCTGCGA TGCAGCGGCG TGAGGCCTGG 300

GCTGGTTGGA GAAGGTCACA ACCCTTCTCT GTTGGTCTGC CTTCTGCTGA AAGACTCGAG 360

AACCAACCAG GGAAGCTGTC CTGGAGGTCC CTGGTCGGAG AGGGACATAG AATCTGTGAC 420

CTCTGACAAC TGTGAAGCCA CCCTGGGCTA CAGAAACCAC AGTCTTCCCA GCAATTATTA 480

CAATTCTTGA ATTCCTTGGG GATTTTTTAC TGCCCTTTCA AAGCACTTAA GTGTTAGATC 540

TAACGTGTTC CAGTGTCTGT CTGAGGTGAC TTAAAAAATC AGAACAAAAC TTCTATTATC 600

CAGAGTCATG GGAGAGTACA CCCTTTCCAG GAATAATGTT TTGGGAAACA CTGAAATGAA 660

ATCTTCCCAG TATTATAAAT TGTGTATTTA AAAAAAAGAA ACTTTTCTGA ATGCCTACTG 720

GCGGTGTATA CCAGGCAGTG TGCCAGTTTA AAAAGATGAA AAAGAATAAA AACTTTTGAG 780

GAACAAAAAA AAAAAAAAAA AAATT 805

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

GGCACGAGAG GCGCCAGTCG CAGGTGTGCT GCTGAGGCGT GAGAATGGCG TCCCGCGGCC 60

GGCGTCCGGA GCATGGCGGA CCCCCAGAGC TGTTTTATGA CGAGACAGAA GCCCGGAAAT 120

ACGTTCGCAA CTCACGGATG ATTGATATCC AGACCAGGAT GGCTGGGCGA GCATTGGAGC 180

TTCTTTATCT GCCAGAGAAT AAGCCCTGTT ACCTGCTGGA TATTGGCTGT GGCACTGGGC 240

TGAGTGGAAG TTATCTGTCA GATGAAGGGC ACTATTGGGT GGGCCTGGAT ATCAGCCCTG 300

CCATGCTGGA TGAGGCTGTG GACCGAGAGA TAGAGGGAGA CCTGCTGCTG GGGGATATGG 360

GCCAGGGCAT CCCATTCAAG CCAGGCACAT TTGATGGTTG CATCAGCATT TCTGCTGTGC 420

AGTGGCTCTG TAATGCTAAC AAGAAGTCTG AAAACCCTGC CAAGCGCCTG TACTGCTTTT 480

TTGCTTCTCT TTTTTCTGTT CTCGTCCGGG GATCCCGAGC TGTCCTGCAG CTGTACCCTG 540

AGAACTCAGA GCAGTTGGAG CTGATCACAA CCCAGGCCAC AAAGGCAGGC TTCTCCGGTG 600

GCATGGTGGT AGACTACCCT AACAGTGCCA AAGCAAAGAA ATTCTACCTC TGCTTGTTTT 660

CTGGGCCTTC GACCTTTATA CCAGAGGGGC TGAGTGAAAA TCAGGATGAA GTTGAACCCA 720

GGGAGTCTGT GTTCACCAAT GAGAGGTTCC CATTAAGGAT GTCGAGGCGG GGAATGGTGA 780

GGAAGAGTCG GGCATGGGTG CTGGAGAAGA AGGAGCGGCA CAGGCGCCAG GGCAGGGAAG 840

TCAGACCTGA CACCCAGTAC ACCGGCCGCA AGCGCAAGCC CCGCTTCTAA GTCACCACGC 900

GGTTCTGGAA AGGCACTTGC CTCTGCACTT TTCTATATTG TTCAGCTGAC AAAGTAGTAT 960

TTTAGAAAAG TTCTAAAGTT ATAAAAATGT TTTCTGCAGT AAAAAAAAAG TTCTCTGGGC 1020

CGGGCGTGGT GGCTCACANC TGTAATCCCA GCACCTTGGG AGGCTGAGGT GGGAGGATCA 1080

TTTGAGGCCA GGAGTTTGAG ACCTGCCTGG GCAACATAAT GAAACTTCCT TTCCAGGGAG 1140

AAAAAAAAAA AAAAAAAAAA ACTCGA 1166

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 586 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

AGAGCGGACG AAGCTGGATA ACAGGGGACC GATGATGTGG CGACCATCAG TTCTGCTGCT 60

TCTGTTGCTA CTGAGGCACG GGGCCCAGGG GAAGCCATCC CCAGACGCAG GCCCTCATGG 120

CCAGGGGAGG GTGCACCAGG CGGCCCCCCT GAGCGACGCT CCCCATGATG ACGCCCACGG 180

GAACTTCCAG TACGACCATG AGGCTTTCCT GGGACGGGAA GTGGCCAAGG AATTCGACCA 240

ACTCACCCCA GAGGAAAGCC AGGCCCGTCT GGGGCGGATC GTGGACCGCA TGGACCGCGC 300

GGGGGACGGC GACGGCTGGG TGTCGCTGGC CGAGCTTCGC GCGTGGATCG CGCACACGCA 360

GCAGCGGCAC ATACGGGACT CGGTGAGCGC GGCCTGGGAC ACGTACGACA CGGACCGCGA 420

CGGGCGTGTG GGTTGGGAGG AGCTGCGCAA CGYCACCTAT GGCCACTASG SGCCCGKTGA 480

AGAATTTCAT GACGTGGAGG ATGCAGAGAC YTACAAAAAG ATGCTGGYTC GGGACGAGCG 540

GCGTTTCCGG GTGGCCGACC AGGATGGGGA CTCGATGGCC ACTCGA 586

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1134 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

ACCCATTGAG CAGAAGGAGG CCAGGTGGGA AAGCTCCTGG GAAGAGCAGC CAGACTGGAC 60

ACTGGGCTGC TTGAGTCCTG AGTCACAATT CAGAATTCCT GGGCTCCCTG GGTGCATTCT 120

ATCATTCCAG TTGAAAGTTT GCTTCCTTCC AGTCATGTGG CTCTTCATTC TACTCTCCTT 180

GGCTCTCATT TCAGATGCCA TGGTCATGGA TGAAAAGGTC AAGAGAAGTT TGTGCTGGAC 240

ACGGCTTCTG CCATCTGCAA CTACAATGCC CAYTACAAGA ATCACCCCAA ATACTGGTGC 300

CGAGGYTATT TCCGTGAYTA CTGCAACATC ATCGCCTTCT CCCCTAACAG CACCAATCAT 360

GTGGCCCTGA AGGACACAGG GAACCAGCTC ATTGTCACTA TGTCCTGCCT GAACAAANAA 420

GACACGGGCT GGTACTGGTG TGGCATCCAR CGGGACTTTG CMAGGGATGA CATGGATTTT 480

ACAGAGCTGA TTGTAACTGA CGACAAAGGA ACCCTGGCCA ATGACTTTTG GTCTGGGAAA 540

GACCTATCAG GCAACAAAAC CAGAAGCTGC AAGGCTCCCA AAGTTGTCCG CAAGCTGACC 600

GCTCCAGGAC GTCCATTCTC ATCATTTGCA TACTGATCAC GGGTTTGGGA ATCATCTCTG 660

TAATCAGTCA TTTGACCAAA AGGAGGAGAA GTCAAAGGAA TAGAAGGGTA GGCAACACTT 720

TGAAGCCCTT CTCGCGTGTC CTGACTCCAA AGGAAATGGC TCCTACTGAA CAGATGTGAC 780

TGAAGWTTTT TTTAATTTAG TTNCATAAAG TGATGNCTAC AACAGAWTAA TCACCCATGA 840

CAACTGGCCC CACACCTCAG AGACTGATTC TGATCTCCCA GGAATTCTGA AGGACCCTCT 900

ATCCTTGACA ACAATCATTT GCAGCCAGGT AGCAACGGCR GTAGTCAGAG GAGCTATGAT 960

AGACCACACC CAAGCAAGGC TGCCCTCAAA TAACATCTCA AGATCTTAGT TCTTATGCAT 1020

TCCATCAGTC AGAAGTGAAG AAGAGGTGGA GAATCTKGAT TGGGGACCAG GAAATCACTT 1080

GTATTTTGTT AGCCAATAAA TTCCTAGCCA GTGTTGAATG AAAAAAAAAA AAAA 1134

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1333 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```
CACTTTAAAG CTCTGCTGAG GGAGTTCGGA GCCCAGGCTT TCAGGCGACC TCTGCCCTCC     60
CTGCCTCTCC TCACCCTCCC TCTCTTCCTG CAGGGCCTGG GAAGGGCTTT GAGGGAGCCT    120
GGGAGCCATG TGAAGAGGGG CACGCCTGGG CTGTCCCACA GTTTAGATCC AGTTGGAGGT    180
TCTCCCTGGC TCCTGCAGGC CTGCGGGGAT CTCTCCCCAC TTCAGGCCTC CGGCAGCTGC    240
CTGCCCTCTT GTCTGTGCTT CAGCCCTGCA CAAAAGCAGC TTGGTGACAC CACTCAGCCA    300
CCCAGAGTAC GTGTTTACAG GCTTTCCAGA TCACCTTCCT GTGGGGTGAA CGTAATGAGG    360
CGGGGCTGGT CCTTGGAATT TCCCCTGGAA AATGGTAACA GACTCCATCC TTGACCCGGG    420
GATGAGCATG AAGGCATTGT CCCAAAGGCA GAGGCCACCG TGGTAGGAAT TCCACCAAGG    480
CCAGAAGGGA AAAAGGAAGA ACCCACCGTG TCTGGCTGTG CGGGCCCTGG GGAGGGTCGT    540
GAGTGCAGCC CCTCTCTACT TCYGTGCCTT TGTAAAACGT GTAGATAACC GCAGTGGTTG    600
GCTGAGCCAA GAACTCTCCT AAATCAGTGG CTTTCTCCCC ACCCCTTGCT GGGGAGTCAT    660
TTTTAAAAAA ATCTGTGGGA TATAAAATTG GCCTCCTGCT GCTTCAGCCT ACCTCTCCCT    720
CTGCTGACTT AATGTCGTGA TTCTGTTTCT TCAGATATTT AAGGCTGTTA GGTTGTGTGA    780
GCCTTGAAGT GTGTGTGTGT GTCCCAGCGA CTGTCCACTG TCCAGGAGAT GCATGTCTTT    840
GTATTGGAGA TATTTCTGTA ACTCATTCTC TTGGTGCTCA CGATTGCCAT GGCCATAGGG    900
CCACAGTGCC GTATCTGCTG CAGACATGAT TGTTTCTTGT TCTAGAGGTT TTCTTGTTTT    960
CGAATCTTGC CTGATGAATC CAGCCAGACC AAGGGGCCTA GATTTGACCT CTGTCCTGGG   1020
CTCCTGGGCC AGGTGCAGGA ACATCTGAGG CCACTCTGCT GGCCACCTCC AGTGGGTGCT   1080
GACCACAGGA TGGGCTTTGT TTACACTCAT TTTCACCCTG ATTCTTGCCC CCACTTTCAT   1140
AAAAGAAACT TCAAAATGCT GACGCTTTGG AGAGTAAGAA AATCAATCTT GGCTGGGCAC   1200
GGTGGCTCCT GCCTGTGATC CTAGCACTTT GGGAGGCTGA AGCTGAAGGA TCACTTGAGC   1260
TCAGGAGTTG GAGACCAACC CTGGCAACAT AACAAGACCC TGTCTCTACA AAAAAAAAAA   1320
AAAAAAAACT CGA                                                      1333
```

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1015 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```
GGCACGAGCG GCACGAGCGG CACGAGGTGA CTTCAAGTGT CGGATCTTTT CAGCCTACAT     60
CAAGGAGGTG GAGGAACGGC CGGCACCCAC CCCGTGGGCT CCAAGATGCC CTTTGGGGAA    120
CTGATGTTCG AATCCAGCAG TAGCTGCGGC TGGGTACATG GCGTCTGTTT CTCAGCCAGC    180
GGGAGCCGCG TGGCCTGGGT AAGCCACGAC AGCACCGTCT GCCTGGCTGA TGCCGACAAG    240
AAGATGGCCG TCGCGACTCT GGCCTCTGAA ACACTACCAC TGCTGGCGCT GACCTTCATC    300
ACAGACAACA GCCTGGTGGC AGCGGGCCAC GACTGCTTCC CGGTGCTGTT CACCTATGAC    360
GCCGCCGCGG GGATGCTGAG CTTCGGCGGG CGGCTGGACG TTCCTAAGCA GAGCTCGCAG    420
CGTGGCTTGA CGGCCCGCGA GCGCTTCCAG AACCTGGACA AGAAGGCGAG CTCCGAGGGT    480
GGCACGGCTG CGGGCGCGGG CCTAGACTCG CTGCACAAGA ACAGCGTCAG CCAGATCTCG    540
GTGCTCAGCG GCGGCAAGGC CAAGTGCTCG CAGTTCTGCA CCACTGGCAT GGATGGCGGC    600
ATGAGTATCT GGGATGTGAA GAGCTTGGAG TCAGCCTTGA AGGACCTCAA GATCAAATGA    660
CCTGTGAGGA ATATGTTGCC TTCATCCTAG CTGCTGGGGA AGCGGGGAGA GGGGTCAGGG    720
AGGCTAATGG TTGCTTTGCT GAATGTTTCT GGGGTACCAA TACGAGTTCC CATAGGGGCT    780
GCTCCCTCAA AAAGGGAGGG GACAGATGGG GAGCTTTTCT TACCTATTCA AGGAATACGT    840
GCCTTTTTCT TAAATGCTTT CATTTATTGA AAAAAAAAAA AAATGCCCCC AAAGCACTAT    900
GCTGGTCATG AACTGCTTCA AAATGTGGAG GTAATAAAAT GCAACTGTGT AAAAAAAAAA    960
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACNC        1015
```

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1076 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
GGCACGAGGG GAAAGCCATG CTCCCAGGAC TCCTTCCTTG CAGCCTTAAA TCGGTCTGTA     60
CGGAAAATTC CGCGCCTTAG AAACCCACGC TTGGGTGTAA CTTATTATTG TTCTTCCTGA    120
CCTACTTCCT GTTTATCACT TCCGGGTTCA TCATTTTGGC ATTTCGGTGA TCGGGTTGGA    180
```

ACTATTGAAG CCCGCTTTCA GGTTCTTTTC CCCATTTTCC CTTTGAAAGG AAGACTTCTG 240

GCTTCTCCTA AATCTCCGTT CTCTGGGTAA GGGGAGTCCA AGCCTCTGTC ATGAGGAACG 300

GAAATGCGAG GGCCTCGGGT GTTACTCTAA AATCCGCCCT CAGCTTGCAC GCCGGAAGCT 360

GCGATTCCTG CAGCGGAAGA GGCGTGATCT GGCCTTCGAC TCGCTATGTC CACTAACAAT 420

ATGTCGGACC CACGGAGGCC GAACAAAGTG CTGAGGTACA AGCCCCCGCC GAGCGAATGT 480

AACCCGGCCT TGGACGACCC GACGCCGGAC TACATGAACC TGCTGGGCAT GATCTTCAGC 540

ATGTGCGGCC TCATGCTTAA GCTGAAGTGG TGTGCTTGGG TCGCTGTCTA CTGCTCCTTC 600

ATCAGCTTTG CCAACTCTCG GAGCTCGGAG GACACGAAGC AAATGATGAG TAGCTTCATG 660

CTGTCCATCT CTGCCGTGGT GATGTCCTAT CTGCAGAATC CTCAGCCCAT GACGCCCCCA 720

TGGTGATACC AGCCTAGAAG GGTCACATTT TGGACCCTGT CTATCCACTA GGCCTGGGCT 780

TTGGCTGCTA AACCTGCTGC CTTCAGCTGC CATCCTGGAC TTCCCTGAAT GAGGCCGTCT 840

CGGTGCCCCC AGCTGGATAG AGGGAACCTG GCCCTTTCCT AGGGAACACC CTAGGCTTAC 900

CCCTCCTGCC TCCCTTCCCC TGCCTGCTGC TGGGGGAGAT GCTGTCCATG TTTCTAGGGG 960

TATTCATTTG CTTTCTCGTT GAAACCTGTT GTTAATAAAG TTTTTCACTC TGAAAAAAAA 1020

AAAAAAAANA RAAAACNCGN GGGGGGGCCC GGAACCCAAT TCSCCGGATA GTGAGT 1076

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1487 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

CCGCTGCTGA TAACTATGGC ATCCCCCGGG CCTGCAGGAA TTCGGCACGG AGCTACGGCG 60

CCGCCTGGCT CCTGCTGNCA CCTGCAGGCT CGTCGCGGGT GGAGCCCACC CAAGACATCA 120

GCATCAGCGA CCAGCTGGGG GGCCAGGACG TGCCCGTGTT CCGGAACCTG TCCCTGCTGG 180

TGGTGGGTGT CGGCGCCGTG TTCTCACTGC TATTCCACCT GGGCACCCGG GAGAGGCGCC 240

GGCCGCATGC GGASGAGCCA GGCGAGCACA CCCCCCTGTT GGCCCCTGCC ACGGCCCAGC 300

CCCTGCTGCT CTGGAAGCAC TGGCTCCGGG AGCSGGCTTT CTACCAGGTG GGCATACTGT 360

ACATGACCAC CAGGCTCATC GTGAACCTGT CCCAGACCTA CATGGCCATG TACCTCACCT 420

ACTCGCTCCA CCTGCCCAAG AAGTTCATCG CGACCATTCC CCTGGTGATG TACCTCAGCG 480

GCTTCTTGTC CTCCTTCCTC ATGAAGCCCA TCAACAAGTG CATTGGGAGG AACATGACCT 540

ACTTCTCAGG CCTCCTGGTG ATCCTGGCCT TTGCCGCCTG GGTGGCGCTG GCGGAGGGAC 600
TGGGTGTGGC CGTGTACGCA GCGGCTGTGC TGCTGGGTGC TGGCTGTGCC ACCATCCTCG 660
TCACCTCGCT GGCCATGACG GCCGACCTCA TCGGTCCCCA CACGAACAGC GGAGCKTTCG 720
TGTACGGCTC CATGAGCTTC TTGGATAAGG TGGCCAATGG GCTGGCAGTC ATGGCCATCC 780
AGAGCCTGCA CCCTTGCCCC TCAGAGCTCT GCTGCAGGGC CTGCGTGAGC TTTTACCACT 840
GGGCGATGGT GGCTGTGACG GGCGGCGTGG GCGTGGCCGC TGCCCTGTGT CTCTGTAGCC 900
TCCTGCTGTG GCCGACCCGC CTGCGACGCT GATGAGACCT GCACGCANTG GCTCACAGCA 960
GCACGATTTG TGACAGCCCG AGGCGGAGAA CACCGAACAC CCAGTGAAGG TGAGGGGATC 1020
AGCACGGCGC GGCCACCCAC GCACCCACGC GCTGGAATGA GACTCAGCCA CAAGGAGGTG 1080
CGAAGCTCTG ACCCAGGCCA CAGTGCGGAT GCACCTTGAG GATGTCACGC TCAGTGAGAG 1140
ACACCAGACA CAGAAGGGTA CGCTGTGATC CCACTTCTAT GAAATGTCCA GGACAGACCA 1200
ATCCACAGAA TCAGGGAGAG GATTCGTGGG TGCCGGGACT GGGGAGGGGG ACCTGGGGGT 1260
GACTAGGTGA CATAATGGGG ACAGGGCTGC CTTCTGGGTG ATGAGAATGT TCTGGAATCA 1320
GATGGGATGG CTGCACGGCG TGGTGAAGGT ACTGAACGCC ACCTCACTGT AAGACGGTAG 1380
ATTTTGTATT TTACCACAAT AAACAAAACA AAACAAAACC AAAAAAAAAA AAAAAAAAAA 1440
AAAAAAAAGG AATTCGATAT CAAGCTTATC GATACCGTCG ACCTCGA 1487

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

GGCACGAGTG CGCANGCGTG GGGCTCTCTC CTTGTCAGTC GGCGCCGCGT GCGGGCTGGT 60
GGCTCTGTGG CAGCGGCGGC GGCAGGACTC CGGCACTATG AGCGGCTTCA GCACCGAGGA 120
GCGCGCCGCG CCTTCTCCCT GGAGTACCGA GTCTTCCTCA AAAATGAGAA AGGACAATAT 180
ATATCTCCAT TTCATGATAT TCCAATTTAT GCAGATAAGG ATGTGTTTCA CATGGTAGTT 240
GAAGTACCAC GCTGGTCTAA TGCAAAAATG GAGATTGCTA CAAAGGACCC TTTAAACCCT 300
ATTAAACAAG ATGTGAAAAA AGGAAAACTT CGCTATGTTG CGAATTTGTT CCCGTATAAA 360
GGATATATCT GGAACTATGG TGCCATCCCT CAGACTTGGG AAGACCCAGG GCACAATGAT 420

AAACATACTG GCTGTTGTGG TGACAATGAC CCAATTGATG TGTGTGAAAT TGGAAGCAAG 480

GTATGTGCAA GAGGTGAAAT AATTGGCGTG AAAGTTCTAG GCATATTGGC TATGATTGAC 540

GAAGGGGAAA CCGACTGGAA AGTCATTGCC ATTAATGTGG ATGATCCTGA TGCAGCCAAT 600

TATAATGATA TCAATGATGT CAAACGGCTG AAACCTGGCT ACTTAGAAGC TACTGTGGAC 660

TGGTTTAGAA GGTATAAGGT TCCTGATGGA AAACCAGAAA ATGAGTTTGC GTTTAATGCA 720

GAATTTAAAG ATAAGGACTT TGCCATTGAT ATTATTAAAA GCACTCATGA CCATTGGAAA 780

GCATTAGTGA CTAAGAAAAC GAATGGAAAA GGAATCAGTT GCATGAATAC AACTTTGTCT 840

GAGAGCCCCT TCAAGTGTGA TCCTGATGCT GCCAGAGCCA TTGTGGATGC TTTACCACCA 900

CCCTGTGAAT CTGCCTGCAC AGTACCAACA GACGTGGATA AGTGGTTCCA TCACCAGAAA 960

AACTAATGAG ATTTCTCTGG AATACAAGCT GATATTGCTA CATCGTGTTC ATCTGGATGT 1020

ATTAGAAGTA AAAGTAGTAG CTTTTCAAAG CTTTAAATTT GTAGAACTCA TCTAACTAAA 1080

GTAAATTCTG CTGTGACTAA TCCAATATAC TCAGAATGTT ATCCATCTAA AGCATTTTTC 1140

ATATCTCAAC TAAGATAACT TTTAGCACAT GCTTAAATAT CAAAGCAGTT GTCATTTGGA 1200

AGTCACTTGT GAATAGATGT GCAAGGGGAG CACATATTGG ATGTATATGT TACCATATGT 1260

TAGGAAATAA AATTATTTTG CTGAAAAAAA AAAAAAAAAA ACCTSGGGGG GGGSCCCGGT 1320

CCCCATTTGG CCCTTTGGGG GGNGGTTTTA 1350

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2527 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

CTCTTGCTAC CTTCCCGGCG CAGAGAACCC CGGCTGCTCA GCGCGCTCCG GGGTCATGGA 60

GATCCCCGGG AGCCTGTGCA AGAAAGTCAA GCTGAGCAAT AACGCGCAGA ACTGGGGAAT 120

GCAGAGAGCA ACCAATGTCA CCTACCAAGC CCATCATGTC AGCAGGAACA AGAGAGGTCA 180

GGTGGTGGGG ACCAGAGGTG GCTTTCGTGG TTGCACAGTT TGGCTAACAG GCTTGTCTGG 240

AGCGGGAAAG ACTACTGTGA GCATGGCCTT GGAGGAGTAC CTGGTTTGTC ATGGTATTCC 300

ATGCTACACT CTGGATGGTG ACAATATTCG TCAAGGTCTC AATAAAAATC TTGGCTTTAG 360

TCCTGAAGAC AGAGAAGAGA ATGTTCGACG CATCGCAGAA GTTGCTAAAC TGTTTGCAGA 420

TGCTGGCTTA GTGTGCATCA CAAGTTTCAT ATCACCTTAC ACTCAGGATC GCAACAATGC 480

AAGGCAAATT CATGAAGGTG CAAGTTTACC GTTTTTTGAA GTATTTGTTG ATGCTCCTCT 540
GCATGTTTGT GAACAGAGGG ATGTCAAAGG ACTCTACAAA AAAGCCCGGG CAGGAGAAAT 600
TAAAGGTTTC ACTGGGATCG ATTCTGAATA TGAAAAGCCA GAGGCCCCTG AGTTGGTGCT 660
GAAAACAGAC TCCTGTGATG TAAATGACTG TGTCCAGCAA GTTGTGGAAC TTCTACAGGA 720
ACGGGATATT GTACCTGTGG ATGCATCTTA TGAAGTAAAA GAACTATATG TGCCAGAAAA 780
TAAACTTCAT TTGGCAAAAA CAGATGCGGA AACATTACCA GCACTGAAAA TTAATAAAGT 840
GGATATGCAG TGGGTGCAGG TTTTGGCAGA AGGTTGGGCA ACCCCATTGA ATGGCTTTAT 900
GAGAGAGAGG GAGTACTTGC AGTGCCTTCA TTTTGATTGT CTTCTGGATG GAGGTGTCAT 960
TAACTTGTCA GTACCTATAG TTCTGACTGC GACTCATGAA GATAAAGAGA GGCTGGACGG 1020
CTGTACAGCA TTTGCTCTGA TGTATGAGGG CCGCCGTGTG GCCATTCTTC GCAATCCAGA 1080
GTTTTTTGAG CACAGGAAAG AGGAGCGCTG TGCCAGACAG TGGGGAACGA CATGCAAGAA 1140
CCACCCCTAT ATTAAGATGG TGATGGAACA AGGAGATTGG CTGATTGGAG GAGATCTTCA 1200
AGTCTTGGAT CGAGTTTATT GGAATGATGG TCTTGATCAG TATCGTCTTA CTCCTACTGA 1260
GCTAAAGCAG AAATTTAAAG ATATGAATGC TGATGCTGTC TTTGCATTTC AACTACGCAA 1320
CCCAGTGCAC AATGGACATG CCCTGTTAAT GCAGGATACC CATAAGCAAC TTCTAGAGAG 1380
GGGCTACCGG CGCCCTGTCC TCCTCCTCCA CCCTCTGGGT GGCTGGACAA AGGATGACGA 1440
TGTTCCTTTG ATGTGGCGTA TGAAGCAGCA TGCTGCAGTG TTGGAGGAAG GAGTTCTGAA 1500
TCCTGAGACG ACAGTGGTGG CCATCTTCCC ATCTCCCATG ATGTATGCTG GACCAACTGA 1560
GGTCCAGTGG CATTGCAGAG CACGGATGGT TGCAGGAGCC AACTTTTACA TTGTTGGACG 1620
AGACCCTGCT GGCATGCCTC ATCCAGAAAC AGGGAAGGAT CTTTATGAGC CAAGTCATGG 1680
TGCCAAAGTG CTGACGATGG CCCCTGGTTT AATCACTTTG GAAATAGTTC CCTTTCGAGT 1740
TGCAGCTTAC AACAAGAAAA AGAAGCGTAT GGACTACTAT GACTCTGAAC ACCATGAAGA 1800
CTTTGAATTT ATTTCAGGAA CACGAATGCG CAAACTTGCT CGAGAAGGCC AGAAACCACC 1860
TGAAGGTTTC ATGGCTCCCA AGGCTTGGAC CGTGCTGACA GAATACTACA AATCCTTGGA 1920
GAAAGCTTAG GCTGTTAACC CAGTCACTCC ACCTTTGACA CATTACTAGT AACAAGAGGG 1980
GACCACATAG TCTCTGTTGG CATTTCTTTG TGGTGTCTGT CTGGACATGC TTCCTAAAAA 2040
CAGACCATTT TCCTTAACTT GCATCAGTTT TGGTCTGCCT TATGAGTTCT GTTTTGAACA 2100
AGTGTAACAC ACTGATGGTT TTAATGTATC TTTTCCACTT ATTATAGTTA TATTCCTACA 2160
ATACAATTTT AAAATTGTCT TTTTATATTA TATTTATGCT TCTGTGTCAT GATTTTTTCA 2220
AGCTGTTATA TTAGTTGTAA CCAGTAGTAT TCACATTAAA TCTTGCTTTT TTTCCCCTTA 2280

```
AAAAAAGAAA AAAATTACCA AACAATAAAC TTGGCTAGAC CTTGTTTTGA GGATTTTACA      2340

AGACCTTTGT AGCGATTAGA TTTTTTTTCT ACATTGAAAA TAGAAACTGC TTCCTTTCTT      2400

CTTTCCAGTC AGCTATTGGT CTTTCCAGCT GTTATAATCT AAAGTATTCT TATGATCTGT      2460

GTAAGCTCTG AATGAACTTC TTTACTCAAT AAAATTAATT TTTTGGCTTC TTAAAAAAAA      2520

AAAAAAA                                                                2527
```

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1098 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

```
CGCATCACAG ACAACCCAGA AGGAAAATGG TTGGGCAGAA CAGCAAGGGG TTCATATGGC        60

TATATTAAAA CAACTGCTGT AGAGATTNNC TATGATTCTT TGAAACTGAA AAAAGACTCT       120

CTTGGTGCCC CTTCAAGACC TATTGAAGAT GACCAAGAAG TATATGATGA TGTTGCAGAG       180

CAGGATGATA TTAGCAGCCA CAGTCAGAGT GGAAGTGGAG GGATATTCCC TCCACCACCA       240

GATGATGACA TTTATGATGG GATTGAAGAG GAAGATGCTG ATGATGGTTT CCCTGCTCCT       300

CCTAAACAAT TGGACATGGG AGATGAAGTT TACGATGATG TGGATACCTC TGATTTCCCT       360

GTTTCATCAG CAGAGATGAG TCAAGGAACT AATGTTGGAA AAGCTAAGAC AGAAGAAAAG       420

GACCTTAAGA AGCTAAAAAA GCAGRAAAAA GAARAAAAAG ACTTCAGGAA AAAATTTAAA       480

TATGATGGTG AAATTAGAGT CCTATATTCA ACTAAAGTTA CAACTTCCAT AACTTCTAAA       540

AAGTGGGGAA CCAGAGATCT ACAGGTAAAA CCTGGTGAAT CTCTAGAAGT TATACAAACC       600

ACAGATGACA CAAAAGTTCT CTGCAGAAAT GAAGAAGGGA AATATGGTTA TGTCCTTCGG       660

AGTTACCTAG CGGACAATGA TGGAGAGATC TATGATGATA TTGCTGATGG CTGCATCTAT       720

GACAATGACT AGCACTCAAC TTTGGTCATT CTGCTGTGTT CATTAGGTGC CAATGTGAAG       780

TCTGGATTTT AATTGGCATG TTATTGGGTA TCMAGAAAAT TAATGCACAR AACCACTTAT       840

TATCATTTGT TATGAAATCC CAATTATCTT TACAAAGTGT TTAAAGTTTG AACATAGAAA       900

ATAATCTCTC TGCTTAATTG TTATCTCAGA AGACTACATT AGTGAGATGT AAGAATTATT       960

AAATATTCCA TTTCCGCTTT GGCTACAATT ATGAAGAAGT TGAAGGTACT TCTTTTAGAC      1020

CACCAGTAAA TAATCCTCCT TCAAAAAATA AAAATAAAAA AAAAAAAAAA AAACTCGAGG      1080
```

GGGGGCCCGG TACCCAAT 1098

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1679 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TCGACCCACG CGTCCGGCGA GATCCCTACC GCAGTAGCCG CCTCTGCCGC CGCGGAGCTT 60

CCCGAACCTC TTCAGCCGCC CGGAGCCGCT CCCGGAGCCC GGCCGTAGAG GCTGCAATCG 120

CAGCCGGGAG CCCGCAGCCC GCGCCCCGAG CCCGCCGCCG CCCTTCGAGG GCGCCCCAGG 180

CCGCGCCATG GTGAAGGTGA CGTTCAACTC CGCTCTGGCC CAGAAGGAGG CCAAGAAGGA 240

CGAGCCCAAG AGCGGCGAGG AGGCGCTCAT CATCCCCCCC GACGCCGTCG CGGTGGACTG 300

CAAGGACCCA GATGATGTGG TACCAGTTGG CCAAAGAAGA GCCTGGTGTT GGTGCATGTG 360

CTTTGGACTA GCATTTATGC TTGCAGGTGT TATTCTAGGA GGAGCATACT TGTACAAATA 420

TTTTGCACTT CAACCAGATG ACGTGTACTA CTGTGGAATA AAGTACATCA AAGATGATGT 480

CATCTTAAAT GAGCCCTCTG CAGATGCCCC AGCTGCTCTC TACCAGACAA TTGAAGAAAA 540

TATTAAAATC TTTGAAGAAG AAGAAGTTGA ATTTATCAGT GTGCCTGTCC CAGAGTTTGC 600

AGATAGTGAT CCTGCCAACA TTGTTCATGA CTTTAACAAG AAACTTACAG CCTATTTAGA 660

TCTTAACCTG GATAAGTGCT ATGTGATCCC TCTGAACACT TCCATTGTTA TGCCACCCAG 720

AAACCTACTG GAGTTACTTA TTAACATCAA GGCTGGAACC TATTTGCCTC AGTCCTATCT 780

GATTCATGAG CACATGGTTA TTACTGATCG CATTGAAAAC ATTGATCACC TGGGTTTCTT 840

TATTTATCGA CTGTGTCATG ACAAGGAAAC TTACAAACTG CAACGCAGAG AAACTATTAA 900

AGGTATTCAG AAACGTGAAG CCAGCAATTG TTTCGCAATT CGGCATTTTG AAAACAAATT 960

TGCCGTGGAA ACTTTAATTT GTTCTTGAAC AGTCAAGAAA AACATTATTG AGGAAAATTA 1020

ATATCACAGC ATAACCCCAC CCTTTACATT TTGTGCAGTG ATTATTTTTT AAAGTCTTCT 1080

TTCATGTAAG TAGCAAACAG GGCTTTACTA TCTTTTCATC TCATTAATTC AATTAAAACC 1140

ATTACCTTAA AATTTTTTTC TTTCGAAGTG TGGTGTCTTT TATATTTGAA TTAGTAACTG 1200

TATGAAGTCA TAGATAATAG TACATGTCAC CTTAGGTAGT AGGAAGAATT ACAATTTCTT 1260

TAAATCATTT ATCTGGATTT TTATGTTTTA TTAGCATTTT CAAGAAGACG GATTATCTAG 1320

AGAATAATCA TATATATGCA TACGTAAAAA TGGACCACAG TGACTTATTT GTAGTTGTTA 1380

GTTGCCCTGC TACCTAGTTT GTTAGTGCAT TTGAGCACAC ATTTTAATTT TCCTCTAATT 1440

AAAATGTGCA GTATTTTCAG TGTCAAATAT ATTTAACTAT TTAGAGAATG ATTTCCACCT 1500

TTATGTTTTA ATATCCTAGG CATCTGCTGT AATAATATTT TAGAAAATGT TTGGAATTTA 1560

AGAAATAACT TGTGTTACTA ATTTGTATAA CCCATATCTG TGCAATGGAA TATAAATATC 1620

ACAAAGTTGT TTAAMWAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAN 1679

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

TTGGCANCNG GGAGAGGGAA AGAGGAGGAA ATGGGGTTTG AGGACCATGG CTTACCTTTC 60

CTGCCTTTGA CCCATCACAC CCCATTTCCT CCTCTTTCCC TCTCCCCGCT GCCAAAAAAA 120

AAAAAAAAGG AAACGTTTAT CATGAATCAA CAGGGTTTCA GTCCTTATCA AAGAGAGATG 180

TGGAAAGAGC TAAAGAAACC ACCCTTTGTT CCCAACTCCA CTTTACCCAT ATTTTATGCA 240

ACACAAACAC TGTCCTTTTG GGTCCCTTTC TTACAGATGG ACCTCTTGAG AAGAATTATC 300

GTATTCCACG TTTTTAGCCC TCAGGTTACC AAGATAAATA TATGTATATA TAACCTTTAT 360

TATTGCTATA TCTTTGTGGA TAATACATTC AGGTGGTGCT GGGTGATTTA TTATAATCTG 420

AACCTAGGTA TATCCTTTGG TCTTCCACAG TCATGTTGAG GTGGGCTCCC TGGTATGGTA 480

AAAAGCCAGG TATAATGTAA CTTCACCCCA GCCTTTGTAC TAAGCTCTTG ATAGTGGATA 540

TACTCTTTTA AGTTTAGCCC CAATATAGGG TAATGGAAAT TTCCTGCCCT CTGGGTTCCC 600

CATTTTTACT ATTAAGAAGA CCAGTGATAA TTTAATAATG CCACCAACTC TGGCTTAGTT 660

AAGTGAGAGT GTGAACTGTG TGGCAAGAGA GCCTCACACC TCACTAGGTG CAGAGAGCCC 720

AGGCCTTATG TTAAAATCAT GCACTTGAAA AGCAAACCTT AATCTGCAAA GACAGCAGCA 780

AGCATTATAC GGTCATCTTG AATGATCCCT TTGAAATTTT TTTTTTGTTT GTTTGTTTAA 840

ATCAAGCCTG AGGCTGGTGA ACAGTAGCTA CACACCCATA TTGTGTGTTC TGTGAATGCT 900

AGCTCTCTTG AATTTGGATA TTGGTTATTT TTTATAGAGT GTAAACCAAG TTTTATATTC 960

TGCAATGCGA ACAGGTACCT ATCTGTTTCT AAATAAAACT GTTTACATTC ATTATGGGGT 1020

ATGTATGACC TTCATTTTCC AAGAAATAGA ACTCTAGCTT AGAATTATGG ATGCTCTAAA 1080

ATGTCAGAAT GGGAACTCTC CTCGAAGTTC TCCCAAACTC AGAGACAGCA CTGCCTTCTC 1140

CTAAATGATT ATTCTTTTCT CCCTGTTTTC TGGTATTTTC TAGGCATCCT TCTCACCACA 1200

GCCATAACCC TTTTTTACTT CCATTAGGCC GTATAACTGG NGGGACNGCT GGTCGGTATA 1260

TAATACTGGT WCCAACAMAG GGGTTCTGGA TGTACACMAG GTTATCTT 1308

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

GGCACAGGAG CGACCCGGGA GAAGGAGGGC CAMGAKGCGG AAGCGGAGGA GTCTCCAGGA 60

GACCCGGGGA CAGCATCGCC CAGGCCCCTG TTTGCAGGCC TTTCAGATAT ATCCATCTCA 120

CAAGACATCC CCGTAGAAGG AGAAATCACC ATTCCTATGA GATCTCGCAT CCGGGAGTTT 180

GACAGCTCCA CATTAAATGA ATCTGTTCGC AATACCATCA TGCGTGATCT AAAAGCTGTT 240

GGGAAAAAAT TCATGCATGT TTTGTACCCA AGGAAAAGTA ATACTCTTTT GAGAGATTGG 300

GATTTGTGGG GCCCTTTGAT CCTTTGTGTG ACACTCGCAT TAATGCTGCA AAGAGACTCT 360

GCAGATAGTG AAAAAGATGG AGGGCCCCAA TTTGCAGAGG TGTTTGTCAT TGTCTGGTTT 420

GGTGCAGTTA CCATCACCCT CAACTCAAAA CTTCTTGGAG GGAACATATC TTTTTTTCAG 480

AGCCTCTGTG TGCTGGGTTA CTGTATACTT CCCTTGACAG TAGCAATGCT GATTTGCCGG 540

CTGGTACTTT TGGCTGATCC AGGACCTGTA AACTTCATGG TTCGGCTTTT TGTGGTGATT 600

GTGATGTTTG CCTGGTCTAT AGTTGCCTCC ACAGCTTTCC TTGCTGATAG CCAGCCTCCA 660

AACCGCAGAG CCCTAGCTGT TTATCCTGTT TTCCTGTTTT ACTTTGTCAT CAGTTGGATG 720

ATTCTCACCT TTACTCCTCA GTAAATCAGG AATGGGAAAT TAAAAACCAG TGAATTGAAA 780

GCACATCTGA AAGATGCAAT TCACCATGGA GCTTTGTCTC TGGCCCTTAT TTGTCTAATT 840

TTGGAGGTAT TTGATAACTG AGTAGGTGAG GAGATTAAAA GGGAGCCATA TAGCACTGTC 900

ACCCCTTATT TGAGGAACTG ATGTTTGAAA GGCTGTTCTT TTCTCTCTTA ATGTCATTTC 960

TTTAAAAATA CATGTGCATA CTACACACAG TATATAATGC CTCCTTAAGG CATGATGGAG 1020

TCACCGTGGT CCATTTGGGT GACAACCAGT GACTTGGGAA GCACATAGAT ACATCTTACA 1080

AGTTGAATAG AGTTGATAAC TATTTTCAGT TTTGAGAATA CCAGTTCAGG TGCAGCTCTT 1140

AAACACATTG CCTTATGACT ATTAGAATAT GCCTCTCTTT TCATAAATAA AAATACATGG 1200

TCTATATCCA TTTTCTTTTA TTTCTCTCTC TTAAGCTTAA AAAGGCAATG AGAGAGGTTA 1260

GGAGTGGGTT CATACACGGA GAATGAGAAA ACATGCATTA ACCAATATTC AGATTTTGAT 1320

CAGGGGAAAT TCTAYACTTG TTGCAAAAAA AAAAAAAAAA AAACTCGAGG GGGGCCCGGT 1380

ACCCAATCGC NGTATATGAT CGNAAACAAT C 1411

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GCTTTGGCTT TTTTTGGCGG ACTGGGGCGC CCTCCGGAAG CGTTTCCAAC TTTCCAGAAG 60

TTTCTCGGGA CGGGCAGGAG GGGGTGGGGA CTGCCATATA TAGATCCCGG GAGCAGGGGA 120

GCGGGCTAAG AGTAGAATCG TGTCGCGGCT CGAGAGCGAG AGTCACGTCC CGGCGCTAGC 180

CAGCCCGACC CAGGCCCACC GTGGTGCACG CAAACCACTT CCTGGCCATG CGCTCCCTCC 240

TGCTTCTCAG CGCCTTCTGC CTCCTGGAGG CGGCCCTGGC CGCCGAGGTG AAGAAACCTG 300

CAGCCGCAGC AGCTCCTGGC ACTGCGGAGA AGTTGAGCCC CAAGGCGGCC ACGCTTGCCG 360

AGCGCANGCC GGCCTGGCCT TCAGCTTGTA CCAGGCCATG GCCAAGGACC AGGCAGTGGA 420

GAACATCCTG GTGTCACCCG TGGTGGTGGC CTCGTCGCTG GGGCTCGTGT CGCTGGGCGG 480

CAAGGCGACC ACGGCGTCGC AGGCCAAGGC AGTGCTGAGC GCCGAGCAGC TGCGCGACGA 540

GGAGGTGCAC GCCGGCCTGG GCGAGCTGCT GCGCTCACTC AGCAACTCGA CGGCGCGCAA 600

CGTGACCTGG AAGCTGGGCA GCCGACTGTA CGGACCCAGC TCAGTGAGCT TCGCTGATGA 660

CTTCGTGCGC ACAGCAAGCA GCACTACAAC TGCGAGCACT CCAAGATCAA CTTCCGCGAC 720

AAGCGCAGNG CGCTGCAGTC CATCAACGAG TGGGCCGCGC AGACCACCGA CGGCAAGCTG 780

CCCGAGGTCA CCAAGGACGT GGAGCGCACG GACGGCGCCC TGCTAGTCAA CGCCATGTTC 840

TTCAAGCCAC ACTGGGATGA GAAATTCCAC CACAAGATGG TGGACAACCG TGGCTTCATG 900

GTGACTCGGT CCTATACYGT GGGTGTCATG ATGATGCACC GGACAGGCCT CTACAACTAC 960

TACGACGACG AGAAGGAAAA GCTGCAAATC GTGGAGATGC CCCTGGCCCA CAAGCTCTCC 1020

AGCCTCATCA TCCTCATGCC CCATCACGTG GAGCCTCTCG AGCGCCTTGA AAAGCTGCTA 1080

ACCAAAGAGC AGCTGAAGAT CTGGATGGGG AAGATGCAGA AGAAGGCTGT TGCCATCTCC 1140

TTGCCCAAGG GTGTGGTGGA GGTGACCCAT GACCTGCAGA AACACCTGGC TGGGCTGGGC 1200

CTGACTGAGG CCATTGACAA GAACAAGGCC GACTTRTCAC GCATGTCAGG CAAGAAGGAC 1260

CTGTACCTGG CCAGCGTGTT CCACGCCACC GCCTTTGAGT TGGACACAGA TGGCAACCCC 1320

TTTGACCAGG ACATCTACGG GCGCGAGGAG CTGCGCANCC CAAGCTGTTC TACGCCGACC 1380

ACCCCTTCAT CTTCCTAGTG CGGGACACCC AAAGCGGCTC CCTGCTATTC ATTGGGCGCC 1440

TGGTCCGGCC TAAGGGTGAC AAGATGCGAG ACGAGTTATA GGGCCTCAGG GTGCACACAG 1500

GATGGCAGGA GGCATCCAAA GGCTCCTGAG ACACATGGGT GCTATTGGGG TTGGGGGGGA 1560

GGTGAGGTAC CAGCCTTGGA TACTCCATGG GGTGGGGGTG GAAAARCAGA CCGGGGTTCC 1620

CGTGTGCCTG AGCGGACCTT CCCAGCTAGA ATTCACTCCA CTTGGACATG GGCCCCAGAT 1680

ACCATGATGC TGAGCCCGGA AACTCCACAT CCTGTGGGAC CTGGGCCATA GTCATTCTGC 1740

CTGCCCTGAA AGTCCCAGAT CAAGCCTGCC TCAATCAGTA TTCATATTTA TAGCCAGGTA 1800

CCTTCTCACC TGTGAGACCA AATTGAGCTA GGGGGGTCAG CCAGCCCTCT TCTGACACTA 1860

AAACACCTCA GCTGCCTCCC CAGCTCTATC CCAACCTCTC CCAACTATAA AACTAGGTGC 1920

TGCAGCCCCT GGGACCAGGC ACCCCCAGAA TGACCTGGCC GCAGTGAGGC GGATTGAGAA 1980

GGAGCTCCCA GGAGGGGCTT CTGGGCAGAC TCTGGTCAAG AAGCATCGTG TCTGGCGTTG 2040

TGGGGATGAA CTTTTTGTTT TGTTTCTTCC TTTTTTAGTT CTTCAAAGAT AGGGAGGGAA 2100

GGGGGAACAT GAGCCTTTGT TGCTATCAAT CCAAGAACTT ATTTGTACAT TTTTTTTTTC 2160

AATAAAACTT TTCCAATGAC AAAAAAAAAA AAAAAAAAAA AAAAAGGGGS GGGCCGCTCC 2220

TAGAGGGATC CCTCCGANGG NGCCCAATCG AAAATN 2256

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 829 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

ATGCGCTCCC TCCTGCTTCT CAGCGCCTTC TGCCTCCTGG AGGCGGCCCT GGCCGCCGAG 60

GTGAAGAAAC CTGCAGCCGC AGCAGCTCCT GGCACTGCGG AGAAGTTGAG CCCCAAGGCG 120

GCCACGCTTG CCGAGCGCAA GCGGCCTGGC CTTCAGCTTG TACCAGGCCA TGGCCAAGGA 180

CCAGGCAGTG GAGAACATCC TGGTGTCACC CGTGGTGGTG GCCTCGTCGC TGGGGCTCGT 240

GTCGCTGGGC GGCAAGGCGA CCACGGCGTC GCAGGCCAAG GCAGTGCTGA GCGCCGAGCA 300

GCTGCGCGAC GAGGAGGTGC ACGCCGGCCT GGGCGAGCTG CTGCGCTCAC TCAGCAACTC 360

CACGGCGCGC AACGTGACCT GGAAGCTGGG CAGCCGACTG TACGGACCCA GCTCAGTGAG 420

CTTCGCTGAT GACTTCGTGC GCAGCAGCAA GCAGCACTAC AACTGCGAGC ACTCCAAGAT 480

CAACTTCCGC GACAAGCGCA GCGCGCTGCA GTCCATCAAC GAGTGGGCCG CGCAGACCAC 540

CGACGGCAAG CTGCCCGAGG TCACCAAGGA CGTGGAGCGC ACGGACGGCG CCCTGTTAGT 600

CAACGCCATG TTCTTCAAGC CACACTGGGA TGAGAAATTC CACCACAAGA TGGTGGACAA 660

CCGTGGCTTC ATGGTGACTC GGTCCTATAC CGTGGGTGTC ATGATGATGC ACCGGACAGG 720

CCTCTACAAC TACTACGACG ACGAGAAGGA AAAGCTGCAA ATCGTGGAGA TGCCCCTGGC 780

CCACAAGCTC TCCAGCCTCA TCATCCTCAT GCCCCATCAC GTGGAGCCT 829

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2223 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

CCTCCGGAAG CGTTTCCAAC TTTCCAGAAG TTTCTCGGGA CGGGCAGGAG GGGGTGGGGA 60

CTGCCATATA TAGATCCCGG GAGCAGGGGA GCGGGCTAAG AGTAGAATCG TGTCGCGGCT 120

CGAGAGCGAG AGTCACGTCC CGGCGCTAGC CAGCCCGACC CAGGCCCACC GTGGTGCACG 180

CAAACCACTT CCTGGCCATG CGCTCCCTCC TGCTTCTCAG CGCCTTCTGC CTCCTGGAGG 240

CGGCCCTGGC CGCCGAGGTG AAGAAACCTG CAGCCGCAGC AGCTCCTGGC ACTGCGGAGA 300

AGTTGAGCCC CAAGGCGGCC ACGCTTGCCG AGCGCAGNCG GCCTGGCCTT CAGCTTGTAC 360

CAGGCCATGG CCAAGGACCA GGCAGTGGAG AACATCCTGG TGTCACCCGT GGTGGTGGCC 420

TCGTCGCTGG GGCTCGTGTC GCTGGGCGGC AAGGCGACCA CGGCGTCGCA GGCCAAGGCA 480

GTGCTGAGCG CCGAGCAGCT GCGCGACGAG GAGGTGCACG CCGGCCTGGG CGAGCTGCTG 540

CGCTCACTCA GCAACTCSAC GGCGCGCAAC GTGACCTGGA AGCTGGGCAG CCGACTGTAC 600

GGACCCAGCT CAGTGAGCTT CGCTGATGAC TTCGTGCGCA CAGCAAGCAG CACTACAACT 660

GCGAGCACTC CAAGATCAAC TTCCGCGACA AGCGCACGCG CTGCAGTCCA TCAACGAGTG 720

GGCCGCGCAG ACCACCGACG GCAAGCTGCC CGAGGTCACC AAGGACGTGG AGCGCACGGA 780

CGGCGCCCTG YTAGTCAACG CCATGTTCTT CAAGCCACAC TGGGATGAGA AATTCCACCA 840

```
CAAGATGGTG GACAACCGTG GCTTCATGGT GACTCGGTCC TATACYGTGG GTGTCATGAT    900
GATGCACCGG ACAGGCCTCT ACAACTACTA CGACGACGAG AAGGAAAAGC TGCAAATCGT    960
GGAGATGCCC CTGGCCCACA AGCTCTCCAG CCTCATCATC CTCATGCCCC ATCACGTGGA   1020
GCCTCTCGAG CGCCTTGAAA AGCTGCTAAC CAAAGAGCAG CTGAAGATCT GGATGGGGAA   1080
GATGCAGAAG AAGGCTGTTG CCATCTCCTT GCCCAAGGGT GTGGTGGAGG TGACCCATGA   1140
CCTGCAGAAA CACCTGGCTG GGCTGGGCCT GACTGAGGCC ATTGACAAGA ACAAGGCCGA   1200
CTTRTCACGC ATGTCAGGCA AGAAGGACCT GTACCTGGCC AGCGTGTTCC ACGCCACCGC   1260
CTTTGAGTTG GACACAGATG GCAACCCCTT TGACCAGGAC ATCTACGGGC GCGAGGAGCT   1320
GCGCASCCCA AGCTGTTCTA CGCCGACCAC CCCTTCATCT TCCTAGTGCG GGACACCCAA   1380
AGCGGCTCCC TGCTATTCAT TGGGCGCCTG GTCCGGCCTA AGGGTGACAA GATGCGAGAC   1440
GAGTTATAGG GCCTCAGGGT GCACACAGGA TGGCAGGAGG CATCCAAAGG CTCCTGAGAC   1500
ACATGGGTGC TATTGGGGTT GGGGGGGAGG TGAGGTACCA GCCTTGGATA CTCCATGGGG   1560
TGGGGGTGGA AAARCAGACC GGGGTTCCCG TGTGCCTGAG CGGACCTTCC CAGCTAGAAT   1620
TCACTCCACT TGGACATGGG CCCCAGATAC CATGATGCTG AGCCCGGAAA CTCCACATCC   1680
TGTGGGACCT GGGCCATAGT CATTCTGCCT GCCCTGAAAG TCCCAGATCA AGCCTGCCTC   1740
AATCAGTATT CATATTTATA GCCAGGTACC TTCTCACCTG TGAGACCAAA TTGAGCTAGG   1800
GGGGTCAGCC AGCCCTCTTC TGACACTAAA ACACCTCAGC TGCCTCCCCA GCTCTATCCC   1860
AACCTCTCCC AACTATAAAA CTAGGTGCTG CAGCCCCTGG GACCAGGCAC CCCCAGAATG   1920
ACCTGGCCGC AGTGAGGCGG ATTGAGAAGG AGCTCCCAGG AGGGGCTTCT GGGCAGACTC   1980
TGGTCAAGAA GCATCGTGTC TGGCGTTGTG GGATGAACT TTTTGTTTTG TTTCTTCCTT   2040
TTTTAGTTCT TCAAAGATAG GGAGGGAAGG GGGAACATGA GCCTTTGTTG CTATCAATCC   2100
AAGAACTTAT TTGTACATTT TTTTTTTCAA TAAAACTTTT CCAATGACAA AAAAAAAAAA   2160
AAAAAAAAAA MWMGGGGSGG GCCGCTCCTA GAGGGATCCC TCCGANGGNG CCCAATCGAA   2220
AAT                                                                 2223
```

(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

Met Lys Lys Gln Ser Lys Arg Cys Leu Trp Lys Pro Pro Gly Ser Leu

```
  1               5                  10                  15

Arg Arg Leu Trp Trp Met Arg Ala Leu Leu Ile Leu Lys Tyr Ile
            20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

```
Met Lys Lys Ser Leu Glu Asn Leu Asn Arg Leu Gln Val Met Leu Leu
  1               5                  10                  15

His Leu Thr Ala Ala Phe Leu Gln Arg Ala His Xaa Ile Leu Thr Thr
            20                  25                  30

Arg Met Ser Leu Gly Phe Gln Ser Pro His Leu Thr Met
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

```
Met His Asn Gln Arg Gln Val Phe Leu Phe His Leu Phe Ser Asn Tyr
  1               5                  10                  15

Leu Leu Ser Ile Asn Ser Val Pro Gly Thr Leu Leu Ala Ala Thr Tyr
            20                  25                  30

Cys Leu Asn Met Thr Tyr Gly
        35
```

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

```
Met Arg Lys Lys Phe Leu Leu Ala Gln Val Phe Leu Ser Leu Ser Val
  1               5                  10                  15

Met Pro Ser Met Pro Val Thr
            20
```

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 110 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

```
Met Val Leu Leu Cys Leu Leu Leu Val Pro Leu Leu Leu Ser Leu Phe
 1               5                  10                  15

Val Leu Gly Leu Phe Leu Trp Phe Leu Lys Arg Glu Arg Gln Glu Glu
            20                  25                  30

Tyr Ile Glu Glu Lys Lys Arg Val Asp Ile Cys Arg Glu Thr Pro Asn
        35                  40                  45

Ile Cys Pro His Ser Gly Glu Asn Thr Glu Tyr Asp Thr Ile Pro His
    50                  55                  60

Thr Asn Arg Thr Ile Leu Lys Glu Asp Pro Ala Asn Thr Val Tyr Ser
65                  70                  75                  80

Thr Val Glu Ile Pro Lys Lys Met Glu Asn Pro His Ser Leu Leu Thr
                85                  90                  95

Met Pro Asp Thr Pro Arg Leu Phe Ala Tyr Glu Asn Val Ile
            100                 105                 110
```

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

```
Met Leu Leu Leu Phe Ile Tyr Phe Tyr Ser His Pro Ala Pro Val Pro
 1               5                  10                  15

Ala Gly Ala Thr Ser Lys Pro Arg Tyr Arg Val Ile Thr Cys Gly Pro
            20                  25                  30

Ala Ser Val Phe Ser Thr Ser Phe Ser His Ser Pro Pro Ala Arg Cys
        35                  40                  45

Leu Gly Arg Leu Glu Gln Met Phe His Phe Gly Leu Ala Ser Gly
    50                  55                  60
```

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

Met Pro Phe Pro Ile Ser Ile Leu Gln Leu Cys Leu Gln Ile Ser Asn
1 5 10 15

Leu Ser Phe Cys Leu Gln Lys Ile Tyr Lys Ile Pro Phe Val
20 25 30

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

Met Ala Ala Ala Cys Arg Ser Val Lys Gly Leu Val Ala Val Ile Thr
1 5 10 15

Gly Gly Ala Ser Gly Leu Gly Leu Ala Thr Ala Asp Asp Leu Trp Gly
20 25 30

Arg Glu Pro Leu Leu Cys Phe Trp Thr Cys Pro Thr Arg Val Gly Arg
35 40 45

Pro Lys Pro Arg Ser
50

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

Met Leu Leu Val Tyr Asp Leu Tyr Leu Xaa Pro Lys Leu Trp Ala Leu
1 5 10 15

Ala Thr Pro Gln Lys Asn Gly Lys Gly Ala Arg Xaa Gly Asp Gly Thr
20 25 30

Pro Ala Gln Ala Phe Trp Asp Phe Trp Ser His Leu Ile Ser Ala Asp
35 40 45

Pro Gln Thr Trp Glu Arg Ala Ala Pro
50 55

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 216 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

Met Arg Leu Ser Ala Leu Leu Ala Leu Ala Ser Lys Val Thr Leu Pro
1 5 10 15

Pro His Tyr Arg Tyr Gly Met Ser Pro Pro Gly Ser Val Ala Asp Lys
20 25 30

Arg Lys Asn Pro Pro Trp Ile Arg Arg Arg Pro Val Val Val Glu Pro
35 40 45

Ile Ser Asp Glu Asp Trp Tyr Leu Phe Cys Gly Asp Thr Val Glu Ile
50 55 60

Leu Glu Gly Lys Asp Ala Gly Lys Gln Gly Lys Val Val Gln Val Ile
65 70 75 80

Arg Gln Arg Asn Trp Val Val Val Gly Gly Leu Asn Thr His Tyr Arg
85 90 95

Tyr Ile Gly Lys Thr Met Asp Tyr Arg Gly Thr Met Ile Pro Ser Glu
100 105 110

Ala Pro Leu Leu His Arg Gln Val Lys Leu Val Asp Pro Met Asp Arg
115 120 125

Lys Pro Thr Glu Ile Glu Trp Arg Phe Thr Glu Ala Gly Glu Arg Val
130 135 140

Arg Val Ser Thr Arg Ser Gly Arg Ile Ile Pro Lys Pro Glu Phe Pro
145 150 155 160

Arg Ala Asp Gly Ile Val Pro Glu Thr Trp Ile Asp Gly Pro Lys Asp
165 170 175

Thr Ser Val Glu Asp Ala Leu Glu Arg Thr Tyr Val Pro Cys Leu Lys
180 185 190

Thr Leu Gln Glu Glu Val Met Glu Ala Met Gly Ile Lys Glu Thr Arg
195 200 205

Lys Tyr Lys Lys Val Tyr Trp Tyr
210 215

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

Met Ser Leu Arg Gln Lys Ser Ser Phe Arg Leu Met Val Met Ser Leu
1 5 10 15

Thr Ile Leu Lys Leu Ser Lys Thr Thr Val Leu Cys Leu Arg Cys Leu
20 25 30

His Ser Leu Lys Leu Thr Trp Arg Asp Gly Ala Arg Cys Ile Asn Ala
35 40 45

Glu

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

Met Ser Gly Ser Phe Ile Leu Cys Leu Ala Leu Val Thr Arg Trp Ser
1 5 10 15

Pro Gln Ala Ser Ser Val Pro Leu Ala Val Tyr Glu Ser Lys Thr Arg
20 25 30

Lys Ser Tyr Arg Ser Gln Arg Asp Arg Asp Gly Lys Asp Arg Ser Gln
35 40 45

Gly Met Gly Leu Ser Leu Leu Val Glu Thr Arg Lys Leu Leu Leu Ser
50 55 60

Ala Asn Gln Gly
65

(2) INFORMATION FOR SEQ ID NO: 137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

Met Cys Phe Arg Phe Phe Leu Phe Cys Ser Arg Ile Leu Leu Lys Leu
1 5 10 15

Phe Phe Leu Leu Phe Pro Ala Ser Ala Phe Pro Leu Ser Thr Arg Ser
20 25 30

Ser Leu Ser Val Asn Glu His Val Val Val Ser Pro Arg Ser Thr Val
35 40 45

Ser Ile Ser Arg
50

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 541 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

```
Met Val Arg Thr Asp Gly His Thr Leu Ser Glu Lys Arg Asn Tyr Gln
1               5                   10                  15

Val Thr Asn Ser Met Phe Gly Ala Ser Arg Lys Lys Phe Val Glu Gly
            20                  25                  30

Val Asp Ser Asp Tyr His Asp Glu Asn Met Tyr Tyr Ser Gln Ser Ser
        35                  40                  45

Met Phe Pro His Arg Ser Glu Lys Asp Met Leu Ala Ser Pro Ser Thr
    50                  55                  60

Ser Gly Gln Leu Ser Gln Phe Gly Ala Ser Leu Tyr Gly Gln Gln Ser
65                  70                  75                  80

Ala Leu Gly Leu Pro Met Arg Gly Met Ser Asn Asn Thr Pro Gln Leu
                85                  90                  95

Asn Arg Ser Leu Ser Gln Gly Thr Gln Leu Pro Ser His Val Thr Pro
            100                 105                 110

Thr Thr Gly Val Pro Thr Met Ser Leu His Thr Pro Pro Ser Pro Ser
        115                 120                 125

Arg Gly Ile Leu Pro Met Asn Pro Xaa Asn Met Met Asn His Ser Gln
    130                 135                 140

Val Gly Gln Gly Ile Gly Ile Pro Ser Arg Thr Asn Ser Met Ser Ser
145                 150                 155                 160

Ser Gly Leu Gly Ser Pro Asn Arg Ser Ser Pro Ser Ile Ile Cys Met
                165                 170                 175

Pro Lys Gln Gln Pro Ser Arg Gln Pro Phe Thr Val Asn Ser Met Ser
            180                 185                 190

Gly Phe Gly Met Asn Arg Asn Gln Ala Phe Gly Met Asn Asn Ser Leu
        195                 200                 205

Ser Ser Asn Ile Phe Asn Gly Thr Asp Gly Ser Glu Asn Val Thr Gly
    210                 215                 220

Leu Asp Leu Ser Asp Phe Pro Ala Leu Ala Asp Arg Asn Arg Arg Glu
225                 230                 235                 240

Gly Ser Gly Asn Pro Thr Pro Leu Ile Asn Pro Leu Ala Gly Arg Ala
                245                 250                 255

Pro Tyr Val Gly Met Val Thr Lys Pro Ala Asn Glu Gln Ser Gln Asp
            260                 265                 270

Phe Ser Ile His Asn Glu Asp Phe Pro Ala Leu Pro Gly Ser Ser Tyr
        275                 280                 285

Lys Asp Pro Thr Ser Ser Asn Asp Asp Ser Lys Ser Asn Leu Asn Thr
    290                 295                 300
```

```
Ser Gly Lys Thr Thr Ser Ser Thr Asp Gly Pro Lys Phe Pro Gly Asp
305                 310                 315                 320

Lys Ser Ser Thr Thr Gln Asn Asn Asn Gln Gln Lys Lys Gly Ile Gln
                325                 330                 335

Val Leu Pro Asp Gly Arg Val Thr Asn Ile Pro Gln Gly Met Val Thr
            340                 345                 350

Asp Gln Phe Gly Met Ile Gly Leu Leu Thr Phe Ile Arg Ala Ala Glu
        355                 360                 365

Thr Asp Pro Gly Met Val His Leu Ala Leu Gly Ser Asp Leu Thr Thr
    370                 375                 380

Leu Gly Leu Asn Leu Asn Ser Pro Glu Asn Leu Tyr Pro Lys Phe Ala
385                 390                 395                 400

Ser Pro Trp Ala Ser Ser Pro Cys Arg Pro Gln Asp Ile Asp Phe His
                405                 410                 415

Val Pro Ser Glu Tyr Leu Thr Asn Ile His Ile Arg Asp Lys Leu Ala
            420                 425                 430

Ala Ile Lys Leu Gly Arg Tyr Gly Glu Asp Leu Leu Phe Tyr Leu Tyr
        435                 440                 445

Tyr Met Asn Gly Gly Asp Val Leu Gln Leu Leu Ala Ala Val Glu Leu
    450                 455                 460

Phe Asn Arg Asp Trp Arg Tyr His Lys Glu Glu Arg Val Trp Ile Thr
465                 470                 475                 480

Arg Ala Pro Gly Met Glu Pro Thr Met Lys Thr Asn Thr Tyr Glu Arg
                485                 490                 495

Gly Thr Tyr Tyr Phe Phe Asp Cys Leu Asn Trp Arg Lys Val Ala Lys
            500                 505                 510

Glu Phe His Leu Glu Tyr Asp Lys Leu Glu Glu Arg Pro His Leu Pro
        515                 520                 525

Ser Thr Phe Asn Tyr Asn Pro Ala Gln Gln Ala Phe Xaa
    530                 535                 540
```

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 58 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

```
Met Ile Cys Pro Gln Cys Pro Leu Ser Leu Leu Cys Leu Ile Ser Ser
1               5                   10                  15

Leu Cys Ser Leu Val Ile Gln Ile Ser Leu Lys Thr Ile Arg Asp Ile
            20                  25                  30
```

Thr Leu Leu Asn Met Val Gly Ile Lys Phe Ser Ile Ser Leu Ser Asn
35 40 45

Lys Ile Asn Ile Asn Ser Arg Thr Trp Xaa
50 55

(2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 202 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

Met Thr Leu Arg Pro Ser Leu Leu Pro Leu His Leu Leu Leu Leu Leu
1 5 10 15

Leu Leu Ser Ala Ala Val Cys Arg Ala Glu Ala Gly Leu Glu Thr Glu
20 25 30

Ser Pro Val Arg Thr Leu Gln Val Glu Thr Leu Val Glu Pro Pro Glu
35 40 45

Pro Cys Ala Glu Pro Ala Ala Phe Gly Asp Thr Leu His Ile His Tyr
50 55 60

Thr Gly Ser Leu Val Asp Gly Arg Ile Ile Asp Thr Ser Leu Thr Arg
65 70 75 80

Asp Pro Leu Val Ile Glu Leu Gly Gln Lys Gln Val Ile Pro Gly Leu
85 90 95

Glu Gln Ser Leu Leu Asp Met Cys Val Gly Glu Lys Arg Arg Ala Ile
100 105 110

Ile Pro Ser His Leu Ala Tyr Gly Lys Arg Gly Phe Pro Pro Ser Val
115 120 125

Pro Ala Asp Ala Val Val Gln Tyr Asp Val Glu Leu Ile Ala Leu Ile
130 135 140

Arg Ala Asn Tyr Trp Leu Lys Leu Val Lys Gly Ile Leu Pro Leu Val
145 150 155 160

Gly Met Ala Met Val Pro Ala Leu Leu Gly Leu Ile Gly Tyr His Leu
165 170 175

Tyr Arg Lys Ala Asn Arg Pro Lys Val Ser Lys Lys Lys Leu Lys Glu
180 185 190

Glu Lys Arg Asn Lys Ser Lys Lys Lys Xaa
195 200

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 217 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

```
Met Phe Leu Arg Leu Tyr Leu Ile Ala Arg Val Met Leu Leu His Ser
 1               5                  10                  15

Lys Leu Phe Thr Asp Ala Ser Ser Arg Ser Ile Gly Ala Leu Asn Lys
            20                  25                  30

Ile Asn Phe Asn Thr Arg Phe Val Met Lys Thr Leu Met Thr Ile Cys
        35                  40                  45

Pro Gly Thr Val Leu Leu Val Phe Ser Ile Ser Leu Trp Ile Ile Ala
    50                  55                  60

Ala Trp Thr Val Arg Val Cys Glu Ser Pro Glu Ser Pro Ala Gln Pro
65                  70                  75                  80

Ser Gly Ser Ser Leu Pro Ala Trp Tyr His Asp Gln Gln Asp Val Thr
                85                  90                  95

Ser Asn Phe Leu Gly Ala Met Trp Leu Ile Ser Ile Thr Phe Leu Ser
            100                 105                 110

Ile Gly Tyr Gly Asp Met Val Pro His Thr Tyr Cys Gly Lys Gly Val
        115                 120                 125

Cys Leu Leu Thr Gly Ile Met Gly Ala Gly Cys Thr Ala Leu Val Val
    130                 135                 140

Ala Val Val Ala Arg Lys Leu Glu Leu Thr Lys Ala Glu Lys His Val
145                 150                 155                 160

His Asn Phe Met Met Asp Thr Gln Leu Thr Lys Arg Ile Lys Asn Ala
                165                 170                 175

Ala Ala Asn Val Leu Arg Glu Thr Trp Leu Ile Tyr Lys His Thr Lys
            180                 185                 190

Leu Leu Lys Lys Ile Asp His Ala Lys Val Arg Lys His Gln Arg Lys
        195                 200                 205

Phe Leu Pro Ser Tyr Pro Pro Val Xaa
    210                 215
```

(2) INFORMATION FOR SEQ ID NO: 142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 102 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

```
Met Ser Asn Thr Thr Val Pro Asn Ala Pro Gln Ala Asn Ser Asp Ser
 1               5                  10                  15
```

Met Val Gly Tyr Val Leu Gly Pro Phe Phe Leu Ile Thr Leu Val Gly
20 25 30

Val Val Val Ala Val Val Met Tyr Val Gln Lys Lys Lys Arg Val Asp
35 40 45

Arg Leu Arg His His Leu Leu Pro Met Tyr Ser Tyr Asp Pro Ala Glu
50 55 60

Glu Leu His Glu Ala Glu Gln Glu Leu Leu Ser Asp Met Gly Asp Pro
65 70 75 80

Lys Val Val His Gly Trp Gln Ser Gly Tyr Gln His Lys Arg Met Pro
85 90 95

Leu Leu Asp Val Lys Thr
100

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 112 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

Met Arg Glu Cys Gln Glu Glu Ser Phe Trp Lys Arg Ala Leu Pro Phe
1 5 10 15

Ser Leu Val Ser Met Leu Val Thr Gln Gly Leu Val Tyr Gln Gly Tyr
20 25 30

Leu Ala Ala Asn Ser Arg Phe Gly Ser Leu Pro Lys Val Ala Leu Ala
35 40 45

Gly Leu Leu Gly Phe Gly Leu Gly Lys Val Ser Tyr Ile Gly Val Cys
50 55 60

Gln Ser Lys Phe His Phe Phe Glu Asp Gln Leu Arg Gly Ala Gly Phe
65 70 75 80

Gly Pro Gln His Asn Arg His Cys Leu Leu Thr Cys Glu Glu Cys Lys
85 90 95

Ile Lys His Gly Leu Ser Glu Lys Gly Asp Ser Gln Pro Ser Ala Ser
100 105 110

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

Met Lys Asn Asp Arg Asn Gln Gly Phe Ser Leu Leu Gln Leu Ile Asp
1 5 10 15

Trp Asn Lys Pro
20

(2) INFORMATION FOR SEQ ID NO: 145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

Met Gly Thr Gln Pro Pro Val Val Ala Gly Phe Thr Ile Pro Met Leu
1 5 10 15

Gly Tyr Thr Val Arg Val Leu Thr Phe His Leu Ser Cys Ser
20 25 30

(2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

Met Lys Ile Pro Val Leu Pro Ala Val Val Leu Leu Ser Leu Leu Val
1 5 10 15

Leu His Ser Ala Gln Gly Ala Thr Leu Gly Gly Pro Glu Glu Glu Ser
20 25 30

Thr Ile Glu Asn Tyr Ala Ser Arg Pro Glu Ala Phe Asn Thr Pro Phe
35 40 45

Leu Asn Ile Asp Lys Leu Arg Ser Ala Phe Lys Ala Asp Glu Phe Leu
50 55 60

Asn Trp His Ala Leu Phe Glu Ser Ile Lys Arg Lys Leu Pro Phe Leu
65 70 75 80

Asn Trp Asp Ala Phe Pro Lys Leu Lys Gly Leu Arg Ser Ala Thr Pro
85 90 95

Asp Ala Gln

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

Met Val Trp Gly Leu Leu Leu Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

Met Leu Pro Leu Leu Ser Leu Leu Phe Leu Phe Phe Ser Thr Val Ser
1 5 10 15

Ser Phe Cys Gly Met Pro Leu Arg Ala His Thr Arg Ala Xaa Ala His
20 25 30

Thr Arg Thr Phe Ala Ser Arg
35

(2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 131 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

Met Ile Cys Glu Thr Lys Ala Arg Lys Ser Ser Gly Gln Pro Gly Arg
1 5 10 15

Leu Pro Pro Pro Thr Leu Ala Pro Pro Gln Pro Pro Leu Pro Glu Thr
20 25 30

Ile Glu Arg Pro Val Gly Thr Gly Ala Met Val Ala Arg Ser Ser Asp
35 40 45

Leu Pro Tyr Leu Ile Val Gly Val Val Leu Gly Ser Ile Val Leu Ile
50 55 60

Ile Val Thr Phe Ile Pro Phe Cys Leu Trp Arg Ala Trp Ser Lys Gln
65 70 75 80

Lys His Thr Thr Asp Leu Gly Phe Pro Arg Ser Ala Leu Pro Pro Ser
85 90 95

Cys Pro Tyr Thr Met Val Pro Leu Gly Gly Leu Pro Gly His Gln Ala
100 105 110

Val Asp Ser Pro Thr Ser Val Ala Ser Val Asp Gly Pro Val Leu Met

115 120 125

Gly Ser Thr
130

(2) INFORMATION FOR SEQ ID NO: 150:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

Met Gly Ala Pro Ser Leu Thr Met Leu Leu Leu Leu Lys Val Gln Pro
1 5 10 15

Arg Arg Thr Gln Ala Phe Asp Ala His Trp Val Gly Leu Pro Leu Leu
20 25 30

(2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

Met Cys Leu Ile Phe Leu Leu Leu Leu Leu Leu Ser Phe Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

His Pro His Gln Asp Ser Gln Pro
1 5

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

Met Asn Thr Ser Tyr Ile Leu Arg Leu Thr Val Val Val Ser Val Val
1 5 10 15

Ile Tyr Leu Ala Ile His Pro Leu Leu Ser Phe Ser Leu Glu Ser Pro
20 25 30

Leu Leu Val Pro Trp Arg Asp Cys Cys Gln Asn Ile Trp Lys Ser Gly
35 40 45

Ser Val Trp Tyr Lys Arg Trp Thr Leu Pro His Met Glu Val Cys Cys
50 55 60

Gln Asp Leu His
65

(2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

Met Leu Lys Ile Phe Lys Glu Trp Glu Asn Leu Asn Leu Ile Leu Thr
1 5 10 15

Ser Ile Arg Ile Leu Glu Arg Gln Asn Met
20 25

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 195 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

Met Asp Cys Glu Val Asn Asn Gly Ser Ser Leu Arg Asp Glu Cys Ile
1 5 10 15

Thr Asn Leu Leu Val Phe Gly Phe Leu Gln Ser Cys Ser Asp Asn Ser
20 25 30

Phe Arg Arg Glu Leu Asp Ala Leu Gly His Glu Leu Pro Val Leu Ala
35 40 45

Pro Gln Trp Glu Gly Tyr Asp Glu Leu Gln Thr Asp Gly Asn Arg Ser
50 55 60

Ser His Ser Arg Leu Gly Arg Ile Glu Ala Asp Ser Glu Ser Gln Glu
65 70 75 80

Asp Ile Ile Arg Asn Ile Ala Arg His Leu Ala Gln Val Gly Asp Ser
85 90 95

Met Asp Arg Ser Ile Pro Pro Gly Leu Val Asn Gly Leu Ala Leu Gln

100 105 110

Leu Arg Asn Thr Ser Arg Ser Glu Glu Asp Arg Asn Arg Asp Leu Ala
115 120 125

Thr Ala Leu Glu Gln Leu Leu Gln Ala Tyr Pro Arg Asp Met Glu Lys
130 135 140

Glu Lys Thr Met Leu Val Leu Ala Leu Leu Leu Ala Lys Lys Val Ala
145 150 155 160

Ser His Thr Pro Ser Leu Leu Arg Asp Val Phe His Thr Thr Val Asn
165 170 175

Phe Ile Asn Gln Asn Leu Arg Thr Tyr Val Arg Ser Leu Ala Arg Asn
180 185 190

Gly Met Asp
195

(2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 91 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

Met Ser Leu Ser Leu Val Ser Val Ser Val Gly Pro Ser Thr Leu Ala
1 5 10 15

Cys Ser Phe Leu Arg Pro Lys Ala Arg Pro Ser Lys Arg Ser Pro Arg
20 25 30

Asn Tyr Thr Asp Ser Thr Ser Pro Gly Gly Pro Arg Ala Pro Arg Gly
35 40 45

Gly Ala Trp Arg Leu Ser Ser Gln Gln Asn Ser Ser Pro Lys Gly Val
50 55 60

Ala Val Ala Lys Ala Ser Tyr Arg Pro Val Leu Cys Phe Leu Pro Gly
65 70 75 80

Pro Trp Ser Ser Xaa Pro Xaa Ala Phe Leu Ile
85 90

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

Met Gly Thr Leu Ser Ala Glu Cys Ser Gly Pro Ala Thr Leu Gly Leu
1 5 10 15

Cys Leu Val Val Pro Trp Asn Ser Ser Gly Leu Ser Gln Pro Pro
20 25 30

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 91 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

Met Lys Phe Leu Ala Val Leu Val Leu Leu Gly Val Ser Ile Phe Leu
1 5 10 15

Val Ser Ala Gln Asn Pro Thr Thr Ala Ala Pro Ala Asp Thr Tyr Pro
20 25 30

Ala Thr Gly Pro Ala Asp Asp Glu Ala Pro Asp Ala Glu Thr Thr Ala
35 40 45

Ala Ala Thr Thr Ala Thr Thr Ala Ala Pro Thr Thr Ala Thr Thr Ala
50 55 60

Ala Ser Thr Thr Ala Arg Lys Asp Ile Pro Val Leu Pro Lys Trp Val
65 70 75 80

Gly Asp Leu Pro Asn Gly Arg Val Cys Pro Xaa
85 90

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

Met Ile Ile Ser Leu Phe Ile Tyr Ile Phe Leu Thr Cys Ser Asn Thr
1 5 10 15

Ser Pro Ser Tyr Gln Gly Thr Gln Leu Gly Leu Gly Leu Pro Ser Ala
20 25 30

Gln Trp Trp Pro Leu Thr Gly Arg Arg Met Gln Cys Cys Arg Leu Phe
35 40 45

Cys Phe Leu Leu Gln Asn Cys Leu Phe Pro Phe Pro Leu His Leu Ile
50 55 60

Gln His Asp Pro Cys Glu Leu Val Leu Thr Ile Ser Trp Asp Trp Ala
65 70 75 80

Glu Ala Gly Ala Ser Leu Tyr Ser Pro
85

(2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

```
Met Ser Ser Ala Ala Ala Asp His Trp Ala Trp Leu Leu Val Leu Ser
 1               5                  10                  15

Phe Val Phe Gly Cys Asn Val Leu Arg Ile Leu Leu Pro Ser Phe Ser
            20                  25                  30

Ser Phe Met Ser Arg Val Leu Gln Lys Asp Ala Glu Gln Glu Ser Gln
        35                  40                  45

Met Arg Ala Glu Ile Gln Asp Met Lys Gln Glu Leu Ser Thr Val Asn
    50                  55                  60

Met Met Asp Glu Phe Ala Arg Tyr Ala Arg Leu Glu Arg Lys Ile Asn
65                  70                  75                  80

Lys Met Thr Asp Lys Leu Lys Thr His Val Lys Ala Arg Thr Ala Gln
                85                  90                  95

Leu Ala Lys Ile Lys Trp Val Ile Ser Val Ala Phe Tyr Val Leu Gln
            100                 105                 110

Ala Ala Leu Met Ile Ser Leu Ile Trp Lys Tyr Tyr Ser Val Pro Val
        115                 120                 125

Ala Val Val Pro Ser Lys Trp Ile Thr Pro Leu Asp Arg Leu Val Ala
    130                 135                 140

Phe Pro Thr Arg Val Ala Gly Gly Val Gly Ile Thr Cys Trp Ile Leu
145                 150                 155                 160

Val Cys Asn Lys Val Val Ala Ile Val Leu His Pro Phe Ser
                165                 170
```

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

```
Met Gly Lys Leu Ile Asn Ile Val Ile Arg Lys Pro Leu Leu Leu Leu
 1               5                  10                  15

Leu Val Gln Cys Glu Asn Cys Cys Arg Lys Asn Met Leu Tyr Asn Ile
            20                  25                  30

Phe Leu Asn Ile His Asn Ile His Lys Phe Ser Asn His
```

35 40 45

(2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

Met Val Ala Ser Thr Leu Val Thr Asn Leu Phe Gly Val Ala Phe Ala
1 5 10 15

Thr Thr Ala Ala Thr Arg Ala
20

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

Met Leu Met Ala Pro Val Val Cys Leu Ser Phe Ser Pro Cys Pro Ala
1 5 10 15

Asp Thr Ser Leu Thr Gly Asp Gly Leu Lys Ala Gly Leu Glu Arg Gly
20 25 30

Xaa Ala Leu Val Thr Leu Phe Asp Ser Val Thr His Phe Leu Ala His
35 40 45

Thr Leu Phe Glu Leu Leu Asp Phe Gln Leu Ala Phe Leu Arg Ser Gly
50 55 60

Lys Gln Thr Ala Pro His
65 70

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 323 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

Met Leu Leu Leu Leu Leu Leu Leu Gly Ser Gly Gln Gly Pro Gln Gln
1 5 10 15

Val Gly Ala Gly Gln Thr Phe Glu Tyr Leu Lys Arg Glu His Ser Leu
20 25 30

Ser Lys Pro Tyr Gln Gly Val Gly Thr Gly Ser Ser Ser Leu Trp Asn

35 40 45

Leu Met Gly Asn Ala Met Val Met Thr Gln Tyr Ile Arg Leu Thr Pro
50 55 60

Asp Met Gln Ser Lys Gln Gly Ala Leu Trp Asn Arg Val Pro Cys Phe
65 70 75 80

Leu Arg Asp Trp Glu Leu Gln Val His Phe Lys Ile His Gly Gln Gly
85 90 95

Lys Lys Asn Leu His Gly Asp Gly Leu Ala Ile Trp Tyr Thr Arg Asn
100 105 110

Arg Met Gln Pro Gly Pro Val Phe Gly Asn Met Asp Lys Phe Val Gly
115 120 125

Leu Gly Val Phe Val Asp Thr Tyr Pro Asn Glu Glu Lys Gln Gln Glu
130 135 140

Arg Val Phe Pro Tyr Ile Ser Ala Met Val Asn Asn Gly Ser Leu Ser
145 150 155 160

Tyr Asp His Glu Arg Asp Gly Arg Pro Thr Glu Leu Gly Gly Cys Thr
165 170 175

Ala Ile Val Arg Asn Leu His Tyr Asp Thr Phe Leu Val Ile Arg Tyr
180 185 190

Val Lys Arg His Leu Thr Ile Met Met Asp Ile Asp Gly Lys His Glu
195 200 205

Trp Arg Asp Cys Ile Glu Val Pro Gly Val Arg Leu Pro Arg Gly Tyr
210 215 220

Tyr Phe Gly Thr Ser Ser Ile Thr Gly Asp Leu Ser Asp Asn His Asp
225 230 235 240

Val Ile Ser Leu Lys Leu Phe Glu Leu Thr Val Glu Arg Thr Pro Glu
245 250 255

Glu Glu Lys Leu His Arg Asp Val Phe Leu Pro Ser Val Asp Asn Met
260 265 270

Lys Leu Pro Glu Met Thr Ala Pro Leu Pro Pro Leu Ser Gly Leu Ala
275 280 285

Leu Phe Leu Ile Val Phe Phe Ser Leu Val Phe Ser Val Phe Ala Ile
290 295 300

Val Ile Gly Ile Ile Leu Tyr Asn Lys Trp Gln Glu Gln Ser Arg Lys
305 310 315 320

Arg Phe Tyr

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 321 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

```
Met Pro Ser Glu Tyr Thr Tyr Val Lys Leu Arg Ser Asp Cys Ser Arg
 1               5                  10                  15

Pro Ser Leu Gln Trp Tyr Thr Arg Ala Gln Ser Lys Met Arg Arg Pro
            20                  25                  30

Ser Leu Leu Leu Lys Asp Ile Leu Lys Cys Thr Leu Leu Val Phe Gly
        35                  40                  45

Val Trp Ile Leu Tyr Ile Leu Lys Leu Asn Tyr Thr Thr Glu Glu Cys
    50                  55                  60

Asp Met Lys Lys Met His Tyr Val Asp Pro Asp His Val Lys Arg Ala
65                  70                  75                  80

Gln Lys Tyr Ala Gln Gln Val Leu Gln Lys Glu Cys Arg Pro Lys Phe
                85                  90                  95

Ala Lys Thr Ser Met Ala Leu Leu Phe Glu His Arg Tyr Ser Val Asp
            100                 105                 110

Leu Leu Pro Phe Val Gln Lys Xaa Pro Lys Asp Ser Glu Ala Glu Ser
        115                 120                 125

Lys Tyr Asp Pro Pro Phe Gly Phe Arg Lys Phe Ser Ser Lys Val Gln
    130                 135                 140

Thr Leu Leu Glu Leu Leu Pro Glu His Asp Leu Pro Glu His Leu Lys
145                 150                 155                 160

Ala Lys Thr Cys Arg Arg Cys Val Val Ile Gly Ser Gly Gly Ile Leu
                165                 170                 175

His Gly Leu Glu Leu Gly His Thr Leu Asn Gln Phe Asp Val Val Ile
            180                 185                 190

Arg Leu Asn Ser Ala Pro Val Glu Gly Tyr Ser Glu His Val Gly Asn
        195                 200                 205

Lys Thr Thr Ile Arg Met Thr Tyr Pro Glu Gly Ala Pro Leu Ser Asp
    210                 215                 220

Leu Glu Tyr Tyr Ser Asn Asp Leu Phe Val Ala Val Leu Phe Lys Ser
225                 230                 235                 240

Val Asp Phe Asn Trp Leu Gln Ala Met Val Lys Lys Glu Thr Leu Pro
                245                 250                 255

Phe Trp Val Arg Leu Phe Phe Trp Lys Gln Val Ala Glu Lys Ile Pro
            260                 265                 270

Leu Gln Pro Lys His Phe Arg Ile Leu Asn Pro Val Ile Ile Lys Glu
        275                 280                 285
```

Thr Ala Phe Xaa His Pro Ser Val Leu Arg Ala Ser Val Lys Val Leu
290 295 300

Gly Ala Glu Ile Arg Thr Ser Pro Gln Ser Val Ser Leu Pro Leu Ser
305 310 315 320

Xaa

(2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

Met Thr Leu Asp Val Gln Thr Val Val Val Phe Ala Val Ile Val Val
1 5 10 15

Leu Leu Leu Val Asn Val Ile Leu Met Phe Phe Leu Gly Thr Arg
20 25 30

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

Met Leu Pro Leu Leu Phe Cys Ala Phe Cys Leu His Lys Leu Gly Pro
1 5 10 15

Leu Leu Phe Leu Tyr Asp Val Leu Met Xaa His Glu Ala Val Met Arg
20 25 30

Thr His Gln Ile Gln Leu Pro Asp Pro Glu Phe Pro Ser Gln Gln Asn
35 40 45

Gln Val Leu Asn Lys Thr Leu Phe Asn Lys Leu Lys Lys Lys Lys Lys
50 55 60

Lys Lys Lys Xaa Xaa Xaa Lys Lys
65 70

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 282 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

```
Met Ala Ser Arg Gly Arg Arg Pro Glu His Gly Gly Pro Pro Glu Leu
1               5                   10                  15

Phe Tyr Asp Glu Thr Glu Ala Arg Lys Tyr Val Arg Asn Ser Arg Met
            20                  25                  30

Ile Asp Ile Gln Thr Arg Met Ala Gly Arg Ala Leu Glu Leu Leu Tyr
        35                  40                  45

Leu Pro Glu Asn Lys Pro Cys Tyr Leu Leu Asp Ile Gly Cys Gly Thr
    50                  55                  60

Gly Leu Ser Gly Ser Tyr Leu Ser Asp Glu Gly His Tyr Trp Val Gly
65                  70                  75                  80

Leu Asp Ile Ser Pro Ala Met Leu Asp Glu Ala Val Asp Arg Glu Ile
                85                  90                  95

Glu Gly Asp Leu Leu Leu Gly Asp Met Gly Gln Gly Ile Pro Phe Lys
            100                 105                 110

Pro Gly Thr Phe Asp Gly Cys Ile Ser Ile Ser Ala Val Gln Trp Leu
        115                 120                 125

Cys Asn Ala Asn Lys Lys Ser Glu Asn Pro Ala Lys Arg Leu Tyr Cys
    130                 135                 140

Phe Phe Ala Ser Leu Phe Ser Val Leu Val Arg Gly Ser Arg Ala Val
145                 150                 155                 160

Leu Gln Leu Tyr Pro Glu Asn Ser Glu Gln Leu Glu Leu Ile Thr Thr
                165                 170                 175

Gln Ala Thr Lys Ala Gly Phe Ser Gly Gly Met Val Val Asp Tyr Pro
            180                 185                 190

Asn Ser Ala Lys Ala Lys Lys Phe Tyr Leu Cys Leu Phe Ser Gly Pro
        195                 200                 205

Ser Thr Phe Ile Pro Glu Gly Leu Ser Glu Asn Gln Asp Glu Val Glu
    210                 215                 220

Pro Arg Glu Ser Val Phe Thr Asn Glu Arg Phe Pro Leu Arg Met Ser
225                 230                 235                 240

Arg Arg Gly Met Val Arg Lys Ser Arg Ala Trp Val Leu Glu Lys Lys
                245                 250                 255

Glu Arg His Arg Arg Gln Gly Arg Glu Val Arg Pro Asp Thr Gln Tyr
            260                 265                 270

Thr Gly Arg Lys Arg Lys Pro Arg Phe Xaa
        275                 280
```

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

Met Leu Gly Lys Thr Lys Phe Gln Ser Tyr Lys Ser Phe Ser Arg Lys
1 5 10 15

Leu Met Val Cys Pro Ser Thr
20

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 328 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

Met Trp Arg Pro Ser Val Leu Leu Leu Leu Leu Leu Leu Arg His Gly
1 5 10 15

Ala Gln Gly Lys Pro Ser Pro Asp Ala Gly Pro His Gly Gln Gly Arg
20 25 30

Val His Gln Ala Ala Pro Leu Ser Asp Ala Pro His Asp Asp Ala His
35 40 45

Gly Asn Phe Gln Tyr Asp His Glu Ala Phe Leu Gly Arg Glu Val Ala
50 55 60

Lys Glu Phe Asp Gln Leu Thr Pro Glu Glu Ser Gln Ala Arg Leu Gly
65 70 75 80

Arg Ile Val Asp Arg Met Asp Arg Ala Gly Asp Gly Asp Gly Trp Val
85 90 95

Ser Leu Ala Glu Leu Arg Ala Trp Ile Ala His Thr Gln Gln Arg His
100 105 110

Ile Arg Asp Ser Val Ser Ala Ala Trp Asp Thr Tyr Asp Thr Asp Arg
115 120 125

Asp Gly Arg Val Gly Trp Glu Glu Leu Arg Asn Ala Thr Tyr Gly His
130 135 140

Tyr Ala Pro Gly Glu Glu Phe His Asp Val Glu Asp Ala Glu Thr Tyr
145 150 155 160

Lys Lys Met Leu Ala Arg Asp Glu Arg Arg Phe Arg Val Ala Asp Gln
165 170 175

Asp Gly Asp Ser Met Ala Thr Arg Glu Glu Leu Thr Ala Phe Leu His
180 185 190

Pro Glu Glu Phe Pro His Met Arg Asp Ile Val Ile Ala Glu Thr Leu
195 200 205

Glu Asp Leu Asp Arg Asn Lys Asp Gly Tyr Val Gln Val Glu Glu Tyr

210 215 220

Ile Ala Asp Leu Tyr Ser Ala Glu Pro Gly Glu Glu Glu Pro Ala Trp
225 230 235 240

Val Gln Thr Glu Arg Gln Gln Phe Arg Asp Phe Arg Asp Leu Asn Lys
245 250 255

Asp Gly His Leu Asp Gly Ser Glu Val Gly His Trp Val Leu Pro Pro
260 265 270

Ala Gln Asp Gln Pro Leu Val Glu Ala Asn His Leu Leu His Glu Ser
275 280 285

Asp Thr Asp Lys Asp Gly Arg Leu Ser Lys Ala Xaa Ile Leu Gly Asn
290 295 300

Trp Asn Met Phe Val Gly Ser Gln Ala Thr Asn Tyr Gly Glu Asp Leu
305 310 315 320

Thr Arg His His Asp Glu Leu Xaa
325

(2) INFORMATION FOR SEQ ID NO: 171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

Met Cys Trp Leu Arg Ala Trp Xaa Gln Ile Xaa Leu Pro Val Phe Xaa
1 5 10 15

Ser Xaa Phe Leu Ile Gln Leu Leu Ile Ser Phe Ser Glu Asn Gly Phe
20 25 30

Ile His Ser Pro Arg Asn Asn Gln Lys Pro Arg Asp Gly Asn Xaa Glu
35 40 45

Glu Cys Ala Val Lys Lys Ser Cys Gln Leu Cys Thr Glu Asp Lys Lys
50 55 60

Tyr Met Met Asn Arg
65

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

Met Trp Leu Phe Ile Leu Leu Ser Leu Ala Leu Ile Ser Asp Ala Met
1 5 10 15

Val Met Asp Glu Lys Val Lys Arg Ser Phe Val Leu Asp Thr Ala Ser
20 25 30

Ala Ile Cys Asn Tyr Asn Ala His Tyr Lys Asn His Pro Lys Tyr Trp
35 40 45

Cys Arg Gly Tyr Phe Arg Asp Tyr Cys Asn Ile Ile Ala Phe Ser Pro
50 55 60

Asn Ser Thr Asn His Val Ala Leu Lys Asp Thr Gly Asn Gln Leu Ile
65 70 75 80

Val Thr Met Ser Cys Leu Asn Lys Glu Asp Thr Gly Trp Tyr Trp Cys
85 90 95

Gly Ile Gln Arg Asp Phe Ala Arg Asp Asp Met Asp Phe Thr Glu Leu
100 105 110

Ile Val Thr Asp Asp Lys Gly Thr Trp Pro Met Thr Leu Val Trp Glu
115 120 125

Arg Leu Ser Gly Thr Lys Pro Glu Ala Ala Arg Leu Pro Lys Leu Ser
130 135 140

Ala Arg Leu Thr Ala Pro Gly Arg Pro Phe Ser Ser Phe Ala Tyr Xaa
145 150 155 160

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 123 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

Met Ala Xaa His Phe Leu Leu Val Ala Leu Gln Ser Val Pro His Cys
1 5 10 15

Pro His Leu Leu Glu Glu Glu His Lys Leu Cys Lys Val Ser His Phe
20 25 30

Ser Gly Val Thr Leu Val Thr Ser Arg Gln Asp Ser Ser Ser Tyr Val
35 40 45

Pro Val Gln Thr Leu Phe Ile His Leu Gly Pro Trp Ala Trp Asp Leu
50 55 60

Xaa Pro Cys Thr Ala Glu Asp Pro Glu Ala Glu Arg Ser Leu Arg Leu
65 70 75 80

Cys His Ser His Leu Ala Arg Xaa Asn Val Ser Pro Ser Gln Ala Ala
85 90 95

Glu Gly Xaa Xaa Xaa Arg Gly Cys Gln His Arg Gly Ser Arg Glu Leu

100 105 110

Thr Phe Leu Ser Ala Glu Asn Glu Ala Gly Ile
115 120

(2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

Met Lys Val Gly Ala Arg Ile Arg Val Lys Met Ser Val Asn Lys Ala
1 5 10 15

His Pro Val Val Ser Thr His Trp Arg Trp Pro Ala Glu Trp Pro Gln
20 25 30

Met Phe Leu His Leu Ala Gln Glu Pro Arg Thr Glu Val Lys Ser Arg
35 40 45

Pro Leu Gly Leu Ala Gly Phe Ile Arg Gln Asp Ser Lys Thr Arg Lys
50 55 60

Pro Leu Glu Gln Glu Thr Ile Met Ser Ala Ala Asp Thr Ala Leu Trp
65 70 75 80

Pro Tyr Gly His Gly Asn Arg Glu His Gln Glu Asn Glu Leu Gln Lys
85 90 95

Tyr Leu Gln Tyr Lys Asp Met His Leu Leu Asp Ser Gly Gln Ser Leu
100 105 110

Gly His Thr His Thr Leu Gln Gly Ser His Asn Leu Thr Ala Leu Asn
115 120 125

Ile

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

Met Ala Tyr His Ser Phe Leu Val Glu Pro Ile Ser Cys His Ala Trp
1 5 10 15

Asn Lys Asp Arg Thr Gln Ile Ala Ile Cys Pro Asn Asn His Glu Val
20 25 30

His Ile Tyr Glu Lys Ser Gly Ala Lys Trp Thr Lys Val His Glu Leu
35 40 45

Lys Glu His Asn Gly Gln Val Thr Gly Ile Asp Trp Ala Pro Glu Ser
50 55 60

Asn Arg Ile Val Thr Cys Gly Thr Asp Arg Asn Ala Tyr Val Trp Thr
65 70 75 80

Leu Lys Gly Arg Thr Trp Lys Pro Thr Leu Val Ile Leu Arg Ile Asn
85 90 95

Arg Ala Ala Arg Cys Val Arg Trp Ala Pro Asn Glu Asn Lys Phe Ala
100 105 110

Val Gly Ser Gly Ser Arg Val Ile Ser Ile Cys Tyr Phe Glu Gln Glu
115 120 125

Asn Asp Trp Trp Val Cys Lys His Ile Lys Lys Pro Ile Arg Ser Thr
130 135 140

Val Leu Ser Leu Asp Trp His Pro Asn Asn Val Leu Leu Ala Ala Gly
145 150 155 160

Ser Cys Asp Phe Lys Cys Arg Ile Phe Ser Ala Tyr Ile Lys Glu Val
165 170 175

Glu Glu Arg Pro Ala Pro Thr Pro Trp Gly Ser Lys Met Pro Phe Gly
180 185 190

Glu Leu Met Phe Glu Ser Ser Ser Ser Cys Gly Trp Val His Gly Val
195 200 205

Cys Phe Ser Ala Ser Gly Ser Arg Val Ala Trp Val Ser His Asp Ser
210 215 220

Thr Val Cys Leu Ala Asp Ala Asp Lys Lys Met Ala Val Ala Thr Leu
225 230 235 240

Ala Ser Glu Thr Leu Pro Leu Leu Ala Leu Thr Phe Ile Thr Asp Asn
245 250 255

Ser Leu Val Ala Ala Gly His Asp Cys Phe Pro Val Leu Phe Thr Tyr
260 265 270

Asp Ala Ala Ala Gly Met Leu Ser Phe Gly Gly Arg Leu Asp Val Pro
275 280 285

Lys Gln Ser Ser Gln Arg Gly Leu Thr Ala Arg Glu Arg Phe Gln Asn
290 295 300

Leu Asp Lys Lys Ala Ser Ser Glu Gly Gly Thr Ala Ala Gly Ala Gly
305 310 315 320

Leu Asp Ser Leu His Lys Asn Ser Val Ser Gln Ile Ser Val Leu Ser
325 330 335

Gly Gly Lys Ala Lys Cys Ser Gln Phe Cys Thr Thr Gly Met Asp Gly
340 345 350

Gly Met Ser Ile Trp Asp Val Lys Ser Leu Glu Ser Ala Leu Lys Asp
355 360 365

Leu Lys Ile Lys
370

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 216 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

Met Trp Ser Ile Gly Ala Gly Ala Leu Gly Ala Ala Ala Leu Ala Leu
1 5 10 15

Leu Leu Ala Asn Thr Asp Val Phe Leu Ser Lys Pro Gln Lys Ala Ala
20 25 30

Leu Glu Tyr Leu Glu Asp Ile Asp Leu Lys Thr Leu Glu Lys Glu Pro
35 40 45

Arg Thr Phe Lys Ala Lys Glu Leu Trp Glu Lys Asn Gly Ala Val Ile
50 55 60

Met Ala Val Arg Arg Pro Gly Cys Phe Leu Cys Arg Glu Glu Ala Ala
65 70 75 80

Asp Leu Ser Ser Leu Lys Ser Met Leu Asp Gln Leu Gly Val Pro Leu
85 90 95

Tyr Ala Val Val Lys Glu His Ile Arg Thr Glu Val Lys Asp Phe Gln
100 105 110

Pro Tyr Phe Lys Gly Glu Ile Phe Leu Asp Glu Lys Lys Lys Phe Tyr
115 120 125

Gly Pro Gln Arg Arg Lys Met Met Phe Met Gly Phe Ile Arg Leu Gly
130 135 140

Val Trp Tyr Asn Phe Phe Arg Ala Trp Asn Gly Gly Phe Ser Gly Asn
145 150 155 160

Leu Glu Gly Glu Gly Phe Ile Leu Gly Gly Val Phe Val Val Gly Ser
165 170 175

Gly Lys Gln Gly Ile Leu Leu Glu His Arg Glu Lys Glu Phe Gly Asp
180 185 190

Lys Val Asn Leu Leu Ser Val Leu Glu Ala Ala Lys Met Ile Lys Pro
195 200 205

Gln Thr Leu Ala Ser Glu Lys Lys
210 215

(2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

Met Lys Pro Val Ser Arg Arg Thr Leu Asp Trp Ile Tyr Ser Val Leu
1 5 10 15

Leu Leu Ala Ile Val Leu Ile Ser Trp Gly Cys Ile Ile Tyr Ala Ser
20 25 30

Met Val Ser Ala Arg Arg Gln Leu Arg Lys Lys Tyr Pro Asp Lys Ile
35 40 45

Phe Gly Thr Asn Glu Asn Leu
50 55

(2) INFORMATION FOR SEQ ID NO: 178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

Met Ala Ala Asn Thr Phe Val Leu Ile Met Gly Ile Pro Thr Ser Ala
1 5 10 15

Asn Ala Xaa Arg Asp Leu Phe
20

(2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

Met Ser Ile Cys His Arg Gly Thr Gly Ile Ala Leu Ser Ala Gly Val
1 5 10 15

Ser Leu Phe Gly Met Ser Ala Leu Leu Leu Pro Gly Asn Phe Glu Ser
20 25 30

Tyr Leu Glu Leu Val Lys Ser Leu Cys Leu Gly Pro Ala Leu Ile His
35 40 45

Thr Ala Lys Phe Ala Leu Val Phe Pro Leu Met Tyr His Thr Trp Asn
50 55 60

Gly Ile Arg His Leu Met Trp Asp Leu Gly Lys Gly Leu Lys Ile Pro
65 70 75 80

Gln Leu Tyr Gln Ser Gly Val Val Val Leu Val Leu Thr Val Leu Ser

85 90 95

Ser Met Gly Leu Ala Ala Met
100

(2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

Met Thr Lys Ala Ser Ser Leu Trp Pro Leu Lys Thr Thr Cys Gln Ile
1 5 10 15

Ser Gly Thr Val Phe Phe Phe Leu Phe Leu Phe Ser Cys Phe Leu Met
20 25 30

Gln Ala Gln Cys Asp Lys Phe Val Gly Trp Asp Phe Phe Phe Phe Leu
35 40 45

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

Met Arg Arg Ala Leu Ile Pro Pro Cys Arg Gly Gly Pro Ser Ala Ser
1 5 10 15

Asp Xaa Cys Cys Ser Cys Ser Pro Ser Gly Phe Ser Ala Gly Arg Gly
20 25 30

Arg Cys Pro Val Gln Gly Cys Leu Arg Pro His Arg Val Gln Leu Leu
35 40 45

Arg Arg Trp Gly Pro Gly Ser Pro Ala Gly Gln Arg Leu Ser Lys Gly
50 55 60

Phe Gln Leu Leu Arg Trp Trp Gly Pro Gly Ser Pro Ala Pro Glu Pro
65 70 75 80

Arg Lys Gly Pro Phe Pro Pro Pro Asp Pro Pro Trp Pro Val Thr Leu
85 90 95

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 95 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

Met Leu Glu Thr Thr Lys His Val Gln Ile Ala Cys Met Leu Leu Leu
1 5 10 15

Thr Cys Gln Ile Phe Leu Pro Ser Ser Leu Ser Pro Ser Phe Ile His
20 25 30

Ser Leu Thr Asp Ser Phe Ile Pro Leu Lys Lys Leu Tyr Val Cys Phe
35 40 45

Val Gln Ser Thr Leu Leu Lys Ala Ala Gly Tyr Lys Ser Ile Ser Glu
50 55 60

Ala Leu Gly Phe Asp Xaa Leu Leu Cys Ser Ser Ala Arg Phe Val Trp
65 70 75 80

Ile Cys His Thr Tyr Ser Arg Pro Leu Val Thr Cys Ala Leu His
85 90 95

(2) INFORMATION FOR SEQ ID NO: 183:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

Met Ser Val Ile Gly Gly Leu Leu Leu Val Val Ala Leu Gly Pro Gly
1 5 10 15

Gly Val Ser Met Asp Glu Lys Lys Lys Glu Trp
20 25

(2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

Met Ser Gly Gly Leu Ser Phe Leu Leu Leu Val
1 5 10

(2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

Met Phe Ala Asp Phe Ile Val Val Thr Ala Thr Val Gln Arg Cys Pro
1   5   10   15

Gly Ser Pro Pro Leu Ser Glu Ile Leu Trp Lys Asp Glu Pro Phe Ala
            20   25   30

Ile Ser Ser His Ala Gly Leu Pro Trp Leu Ser Ser Trp Pro Ala Pro
        35   40   45

Pro Trp Thr Trp Ser Trp Ile Ser Arg Arg Arg Glu His Gly Arg Gly
    50   55   60

Ser
65

(2) INFORMATION FOR SEQ ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

Met Val Glu Ser Val Met Pro Val Val Val Cys Thr Leu Ser Pro Gly
1   5   10   15

Ile Asp Ser Ser Pro Ser
            20

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 132 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

Met Asp Val Leu Phe Val Ala Ile Phe Ala Val Pro Leu Ile Leu Gly
1   5   10   15

Gln Glu Tyr Glu Asp Glu Glu Arg Leu Gly Glu Asp Glu Tyr Tyr Gln
            20   25   30

Val Val Tyr Tyr Tyr Thr Val Thr Pro Ser Tyr Asp Asp Phe Ser Ala
        35   40   45

Asp Phe Thr Ile Asp Tyr Ser Ile Phe Glu Ser Glu Asp Arg Leu Asn
    50   55   60

Arg Leu Asp Lys Asp Ile Thr Glu Ala Ile Glu Thr Thr Ile Ser Leu
65   70   75   80

Glu Thr Ala Arg Ala Asp His Pro Lys Pro Val Thr Val Lys Pro Val
85 90 95

Thr Thr Glu Pro Gln Ser Pro Asp Leu Asn Asp Ala Val Ser Ser Leu
100 105 110

Arg Ser Pro Ile Pro Leu Leu Leu Ser Cys Ala Phe Val Gln Val Gly
115 120 125

Met Tyr Phe Met
130

(2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

Met Pro Cys Gln Pro Gly Gln Val Pro Ser Cys Gln Cys Thr Phe Gly
1 5 10 15

Leu Leu Leu Met Leu Pro Ser Leu Pro Ser Pro Ala Ser Gln Pro Arg
20 25 30

Pro Phe Cys Ser Ser Met Glu Tyr Phe His Gly Cys Ala Ser Pro Ser
35 40 45

Gln Ala Ile Ile Gly Gly Phe Pro Phe Ala Ser Val Ala Leu Ala Asp
50 55 60

Ile Leu Cys Leu Gln
65

(2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

Met Ser Leu Leu Ser Pro Ala Ile Pro Ala Leu Thr Leu Ile Phe Ile
1 5 10 15

Leu Met Phe Phe Ser Phe Pro Phe Arg Ala His Thr Val Val Thr Ile
20 25 30

Val Ala Ser Gly Phe Leu Gly Leu Ser Pro Leu Cys Gly
35 40 45

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

Met Ala Phe Gly Leu Gln Met Phe Ile Gln Arg Lys Phe Pro Tyr Pro
1 5 10 15

Leu Gln Trp Ser Leu Leu Val Ala Val Val Ala Gly Ser Val Val Ser
20 25 30

Tyr Gly Val Thr Arg Val Glu Ser Glu Lys Cys Asn Asn Leu Trp Leu
35 40 45

Phe Leu Glu Thr Gly Gln Leu Pro Lys Asp Arg Ser Thr Asp Gln Arg
50 55 60

Ser
65

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

Met Asn Leu Leu Gly Met Ile Phe Ser Met Cys Gly Leu Met Leu Lys
1 5 10 15

Leu Lys Trp Cys Ala Trp Val Ala Val Tyr Cys Ser Phe Ile Ser Phe
20 25 30

Ala Asn Ser Arg Ser Ser Glu Asp Thr Lys Gln Met Met Ser Ser Phe
35 40 45

Met Xaa
50

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 170 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

Met Leu Leu Asn Val Ala Leu Val Ala Leu Val Leu Leu Gly Ala Tyr
1 5 10 15

Arg Leu Trp Val Arg Trp Gly Arg Arg Gly Leu Gly Ala Gly Ala Gly
20 25 30

Ala Gly Glu Glu Ser Pro Ala Thr Ser Leu Pro Arg Met Lys Lys Arg
35 40 45

Asp Phe Ser Leu Glu Gln Leu Arg Gln Tyr Asp Gly Ser Arg Asn Pro
50 55 60

Arg Ile Leu Leu Ala Val Asn Gly Lys Val Phe Asp Val Thr Lys Gly
65 70 75 80

Ser Lys Phe Tyr Gly Pro Ala Gly Pro Tyr Gly Ile Phe Ala Gly Arg
85 90 95

Asp Ala Ser Arg Gly Leu Ala Thr Phe Cys Leu Asp Lys Asp Ala Leu
100 105 110

Arg Asp Glu Tyr Asp Asp Leu Ser Asp Leu Asn Ala Val Gln Met Glu
115 120 125

Ser Val Arg Glu Trp Glu Met Gln Phe Lys Glu Lys Tyr Asp Tyr Val
130 135 140

Gly Arg Leu Leu Lys Pro Gly Glu Glu Pro Ser Glu Tyr Thr Asp Glu
145 150 155 160

Glu Asp Thr Lys Asp His Asn Lys Gln Asp
165 170

(2) INFORMATION FOR SEQ ID NO: 193:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

Met Thr Tyr Phe Ser Gly Leu Leu Val Ile Leu Ala Phe Ala Ala Trp
1 5 10 15

Val Ala Leu Ala Glu Gly Leu Gly Val Ala Val Tyr Ala Ala Ala Val
20 25 30

Leu Leu Gly Ala Gly Cys Ala Thr Ile Leu Val Thr Ser Leu Ala Met
35 40 45

Thr Ala Asp Leu Ile Gly Pro His Thr Asn Ser Gly Leu Ser Cys Thr
50 55 60

Ala Pro
65

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 92 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

Met Ala Ala Gly Pro Ser Gly Cys Leu Val Pro Ala Phe Gly Leu Arg
1 5 10 15

Leu Leu Leu Ala Thr Val Leu Gln Ala Val Ser Ala Phe Gly Ala Glu
20 25 30

Phe Ser Ser Glu Ala Cys Arg Glu Leu Gly Phe Ser Ser Asn Leu Leu
35 40 45

Cys Ser Ser Cys Asp Leu Leu Gly Gln Phe Asn Leu Leu Gln Leu Asp
50 55 60

Pro Asp Cys Arg Gly Cys Cys Gln Glu Glu Ala Gln Phe Glu Thr Lys
65 70 75 80

Lys Leu Tyr Ala Gly Ala Ile Leu Glu Val Cys Gly
85 90

(2) INFORMATION FOR SEQ ID NO: 195:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 176 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

Met Arg Gly Ser His Leu Arg Leu Leu Pro Tyr Leu Val Ala Ala Asn
1 5 10 15

Pro Val Asn Tyr Gly Arg Pro Tyr Arg Leu Ser Cys Val Glu Ala Phe
20 25 30

Ala Ala Thr Phe Cys Ile Val Gly Phe Pro Asp Leu Ala Val Ile Leu
35 40 45

Leu Arg Lys Phe Lys Trp Gly Lys Gly Phe Leu Asp Leu Asn Arg Gln
50 55 60

Leu Leu Asp Lys Tyr Ala Ala Cys Gly Ser Pro Glu Glu Val Leu Gln
65 70 75 80

Ala Glu Gln Glu Phe Leu Ala Asn Ala Lys Glu Ser Pro Gln Glu Glu
85 90 95

Glu Ile Asp Pro Phe Asp Val Asp Ser Gly Arg Glu Phe Gly Asn Pro
100 105 110

Asn Arg Pro Val Ala Ser Thr Arg Leu Pro Ser Asp Thr Asp Asp Ser
115 120 125

Asp Ala Ser Glu Asp Pro Gly Pro Xaa Ala Glu Arg Gly Gly Ala Ser
130 135 140

Ser Ser Cys Cys Glu Glu Glu Gln Thr Gln Gly Arg Gly Ala Glu Ala
145 150 155 160

Arg Ala Pro Ala Glu Val Trp Lys Gly Ile Lys Lys Arg Gln Arg Asp
165 170 175

(2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

Met Ser Asn Ala Cys Lys Glu Leu Ala Ile Phe Leu Thr Thr Gly Ile
1 5 10 15

Val Val Ser Ala Phe Gly Leu Pro Ile Val Phe Ala Arg Ala His Leu
20 25 30

Ile Glu Trp Gly Ala Cys Ala Leu Val Leu Thr Gly Asn Thr Val Ile
35 40 45

Phe Ala Thr Ile Leu Gly Phe Phe Leu Val Phe Gly Ser Asn Asp Asp
50 55 60

Phe Ser Trp Gln Gln Trp
65 70

(2) INFORMATION FOR SEQ ID NO: 197:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

Met Thr Leu Leu Ile Ile Phe Leu Pro Phe Xaa Phe Thr Thr Xaa Thr
1 5 10 15

Asn Ser Gly Gly Ser Phe Pro Val Arg
20 25

(2) INFORMATION FOR SEQ ID NO: 198:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

Met Lys Gly Glu Leu Leu Pro Phe Leu Phe Leu Thr Val Trp Leu Trp
1 5 10 15

Leu Tyr Lys Leu Xaa Phe Gly Glu Ser Pro Arg Tyr Pro Asn Val Ile
20 25 30

Gly Lys Thr Tyr Phe Phe Phe Trp Thr Asp Gln Ile Ser Arg Glu Ser
35 40 45

Arg Phe Leu Glu Arg Leu Ala Phe Ile Val Ser Glu Asn Cys Leu Ile
50 55 60

Phe Leu Ile His Ala Ile Thr Gly Gln
65 70

(2) INFORMATION FOR SEQ ID NO: 199:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 289 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

Met Ser Gly Phe Ser Thr Glu Glu Arg Ala Ala Pro Phe Ser Leu Glu
1 5 10 15

Tyr Arg Val Phe Leu Lys Asn Glu Lys Gly Gln Tyr Ile Ser Pro Phe
20 25 30

His Asp Ile Pro Ile Tyr Ala Asp Lys Asp Val Phe His Met Val Val
35 40 45

Glu Val Pro Arg Trp Ser Asn Ala Lys Met Glu Ile Ala Thr Lys Asp
50 55 60

Pro Leu Asn Pro Ile Lys Gln Asp Val Lys Lys Gly Lys Leu Arg Tyr
65 70 75 80

Val Ala Asn Leu Phe Pro Tyr Lys Gly Tyr Ile Trp Asn Tyr Gly Ala
85 90 95

Ile Pro Gln Thr Trp Glu Asp Pro Gly His Asn Asp Lys His Thr Gly
100 105 110

Cys Cys Gly Asp Asn Asp Pro Ile Asp Val Cys Glu Ile Gly Ser Lys
115 120 125

Val Cys Ala Arg Gly Glu Ile Ile Gly Val Lys Val Leu Gly Ile Leu
130 135 140

Ala Met Ile Asp Glu Gly Glu Thr Asp Trp Lys Val Ile Ala Ile Asn
145 150 155 160

Val Asp Asp Pro Asp Ala Ala Asn Tyr Asn Asp Ile Asn Asp Val Lys
165 170 175

Arg Leu Lys Pro Gly Tyr Leu Glu Ala Thr Val Asp Trp Phe Arg Arg
180 185 190

Tyr Lys Val Pro Asp Gly Lys Pro Glu Asn Glu Phe Ala Phe Asn Ala
195 200 205

Glu Phe Lys Asp Lys Asp Phe Ala Ile Asp Ile Ile Lys Ser Thr His
    210   215   220

Asp His Trp Lys Ala Leu Val Thr Lys Lys Thr Asn Gly Lys Gly Ile
225   230   235   240

Ser Cys Met Asn Thr Thr Leu Ser Glu Ser Pro Phe Lys Cys Asp Pro
                245   250   255

Asp Ala Ala Arg Ala Ile Val Asp Ala Leu Pro Pro Pro Cys Glu Ser
            260   265   270

Ala Cys Thr Val Pro Thr Asp Val Asp Lys Trp Phe His His Gln Lys
        275   280   285

Asn

(2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 625 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

Met Glu Ile Pro Gly Ser Leu Cys Lys Lys Val Lys Leu Ser Asn Asn
1   5   10   15

Ala Gln Asn Trp Gly Met Gln Arg Ala Thr Asn Val Thr Tyr Gln Ala
            20   25   30

His His Val Ser Arg Asn Lys Arg Gly Gln Val Val Gly Thr Arg Gly
        35   40   45

Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly Leu Ser Gly Ala Gly
    50   55   60

Lys Thr Thr Val Ser Met Ala Leu Glu Glu Tyr Leu Val Cys His Gly
65   70   75   80

Ile Pro Cys Tyr Thr Leu Asp Gly Asp Asn Ile Arg Gln Gly Leu Asn
                85   90   95

Lys Asn Leu Gly Phe Ser Pro Glu Asp Arg Glu Glu Asn Val Arg Arg
            100   105   110

Ile Ala Glu Val Ala Lys Leu Phe Ala Asp Ala Gly Leu Val Cys Ile
        115   120   125

Thr Ser Phe Ile Ser Pro Tyr Thr Gln Asp Arg Asn Asn Ala Arg Gln
    130   135   140

Ile His Glu Gly Ala Ser Leu Pro Phe Phe Glu Val Phe Val Asp Ala
145   150   155   160

Pro Leu His Val Cys Glu Gln Arg Asp Val Lys Gly Leu Tyr Lys Lys

165 170 175

Ala Arg Ala Gly Glu Ile Lys Gly Phe Thr Gly Ile Asp Ser Glu Tyr
180 185 190

Glu Lys Pro Glu Ala Pro Glu Leu Val Leu Lys Thr Asp Ser Cys Asp
195 200 205

Val Asn Asp Cys Val Gln Gln Val Val Glu Leu Leu Gln Glu Arg Asp
210 215 220

Ile Val Pro Val Asp Ala Ser Tyr Glu Val Lys Glu Leu Tyr Val Pro
225 230 235 240

Glu Asn Lys Leu His Leu Ala Lys Thr Asp Ala Glu Thr Leu Pro Ala
245 250 255

Leu Lys Ile Asn Lys Val Asp Met Gln Trp Val Gln Val Leu Ala Glu
260 265 270

Gly Trp Ala Thr Pro Leu Asn Gly Phe Met Arg Glu Arg Glu Tyr Leu
275 280 285

Gln Cys Leu His Phe Asp Cys Leu Leu Asp Gly Gly Val Ile Asn Leu
290 295 300

Ser Val Pro Ile Val Leu Thr Ala Thr His Glu Asp Lys Glu Arg Leu
305 310 315 320

Asp Gly Cys Thr Ala Phe Ala Leu Met Tyr Glu Gly Arg Arg Val Ala
325 330 335

Ile Leu Arg Asn Pro Glu Phe Phe Glu His Arg Lys Glu Glu Arg Cys
340 345 350

Ala Arg Gln Trp Gly Thr Thr Cys Lys Asn His Pro Tyr Ile Lys Met
355 360 365

Val Met Glu Gln Gly Asp Trp Leu Ile Gly Gly Asp Leu Gln Val Leu
370 375 380

Asp Arg Val Tyr Trp Asn Asp Gly Leu Asp Gln Tyr Arg Leu Thr Pro
385 390 395 400

Thr Glu Leu Lys Gln Lys Phe Lys Asp Met Asn Ala Asp Ala Val Phe
405 410 415

Ala Phe Gln Leu Arg Asn Pro Val His Asn Gly His Ala Leu Leu Met
420 425 430

Gln Asp Thr His Lys Gln Leu Leu Glu Arg Gly Tyr Arg Arg Pro Val
435 440 445

Leu Leu Leu His Pro Leu Gly Gly Trp Thr Lys Asp Asp Asp Val Pro
450 455 460

Leu Met Trp Arg Met Lys Gln His Ala Ala Val Leu Glu Glu Gly Val
465 470 475 480

Leu Asn Pro Glu Thr Thr Val Val Ala Ile Phe Pro Ser Pro Met Met

485 490 495

Tyr Ala Gly Pro Thr Glu Val Gln Trp His Cys Arg Ala Arg Met Val
500 505 510

Ala Gly Ala Asn Phe Tyr Ile Val Gly Arg Asp Pro Ala Gly Met Pro
515 520 525

His Pro Glu Thr Gly Lys Asp Leu Tyr Glu Pro Ser His Gly Ala Lys
530 535 540

Val Leu Thr Met Ala Pro Gly Leu Ile Thr Leu Glu Ile Val Pro Phe
545 550 555 560

Arg Val Ala Ala Tyr Asn Lys Lys Lys Lys Arg Met Asp Tyr Tyr Asp
565 570 575

Ser Glu His His Glu Asp Phe Glu Phe Ile Ser Gly Thr Arg Met Arg
580 585 590

Lys Leu Ala Arg Glu Gly Gln Lys Pro Pro Glu Gly Phe Met Ala Pro
595 600 605

Lys Ala Trp Thr Val Leu Thr Glu Tyr Tyr Lys Ser Leu Glu Lys Ala
610 615 620

Xaa
625

(2) INFORMATION FOR SEQ ID NO: 201:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 649 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

Met Ser Ala Ser Gln Asp Leu Glu Pro Lys Pro Leu Phe Pro Lys Pro
1 5 10 15

Ala Phe Gly Gln Lys Pro Pro Leu Ser Thr Glu Asn Ser His Glu Asp
20 25 30

Glu Ser Pro Met Lys Asn Val Ser Ser Ser Lys Gly Ser Pro Ala Pro
35 40 45

Leu Gly Val Arg Ser Lys Ser Gly Pro Leu Lys Pro Ala Arg Glu Asp
50 55 60

Ser Glu Asn Lys Asp His Ala Gly Glu Ile Ser Ser Leu Pro Phe Pro
65 70 75 80

Gly Val Val Leu Lys Pro Ala Ala Ser Arg Gly Gly Pro Gly Leu Ser
85 90 95

Lys Asn Gly Glu Glu Lys Lys Glu Asp Arg Lys Ile Asp Ala Ala Lys
100 105 110

Asn Thr Phe Gln Ser Lys Ile Asn Gln Glu Glu Leu Ala Ser Gly Thr
115 120 125

Pro Pro Ala Arg Phe Pro Lys Ala Pro Ser Lys Leu Thr Val Gly Gly
130 135 140

Pro Trp Gly Gln Ser Gln Glu Lys Glu Lys Gly Asp Lys Asn Ser Ala
145 150 155 160

Thr Pro Lys Gln Lys Pro Leu Pro Pro Leu Phe Thr Leu Gly Pro Pro
165 170 175

Pro Pro Lys Pro Asn Arg Pro Pro Asn Val Asp Leu Thr Lys Phe His
180 185 190

Lys Thr Ser Ser Gly Asn Ser Thr Ser Lys Gly Gln Thr Ser Tyr Ser
195 200 205

Thr Thr Ser Leu Pro Pro Pro Pro Pro Ser His Pro Ala Ser Gln Pro
210 215 220

Pro Leu Pro Ala Ser His Pro Ser Gln Pro Pro Val Pro Ser Leu Pro
225 230 235 240

Pro Arg Asn Ile Lys Pro Pro Phe Asp Leu Lys Ser Pro Val Asn Glu
245 250 255

Asp Asn Gln Asp Gly Val Thr His Ser Asp Gly Ala Gly Asn Leu Asp
260 265 270

Glu Glu Gln Asp Ser Glu Gly Glu Thr Tyr Glu Asp Ile Glu Ala Ser
275 280 285

Lys Glu Arg Glu Lys Lys Arg Glu Lys Glu Glu Lys Lys Arg Leu Glu
290 295 300

Leu Glu Lys Lys Glu Gln Lys Glu Lys Glu Lys Lys Glu Gln Glu Ile
305 310 315 320

Lys Lys Lys Phe Lys Leu Thr Gly Pro Ile Gln Val Ile His Leu Ala
325 330 335

Lys Ala Cys Cys Asp Val Lys Gly Gly Lys Asn Glu Leu Ser Phe Lys
340 345 350

Gln Gly Glu Gln Ile Glu Ile Ile Arg Ile Thr Asp Asn Pro Glu Gly
355 360 365

Lys Trp Leu Gly Arg Thr Ala Arg Gly Ser Tyr Gly Tyr Ile Lys Thr
370 375 380

Thr Ala Val Glu Ile Asp Tyr Asp Ser Leu Lys Leu Lys Lys Asp Ser
385 390 395 400

Leu Gly Ala Pro Ser Arg Pro Ile Glu Asp Asp Gln Glu Val Tyr Asp
405 410 415

Asp Val Ala Glu Gln Asp Asp Ile Ser Ser His Ser Gln Ser Gly Ser
420 425 430

Gly Gly Ile Phe Pro Pro Pro Pro Asp Asp Asp Ile Tyr Asp Gly Ile
435 440 445

Glu Glu Glu Asp Ala Asp Asp Gly Ser Thr Leu Gln Val Gln Glu Lys
450 455 460

Ser Asn Thr Trp Ser Trp Gly Ile Leu Lys Met Leu Lys Gly Lys Asp
465 470 475 480

Asp Arg Lys Lys Ser Ile Arg Glu Lys Pro Lys Val Ser Asp Ser Asp
485 490 495

Asn Asn Glu Gly Ser Ser Phe Pro Ala Pro Pro Lys Gln Leu Asp Met
500 505 510

Gly Asp Glu Val Tyr Asp Asp Val Asp Thr Ser Asp Phe Pro Val Ser
515 520 525

Ser Ala Glu Met Ser Gln Gly Thr Asn Val Gly Lys Ala Lys Thr Glu
530 535 540

Glu Lys Asp Leu Lys Lys Leu Lys Lys Gln Xaa Lys Xaa Xaa Lys Asp
545 550 555 560

Phe Arg Lys Lys Phe Lys Tyr Asp Gly Glu Ile Arg Val Leu Tyr Ser
565 570 575

Thr Lys Val Thr Thr Ser Ile Thr Ser Lys Lys Trp Gly Thr Arg Asp
580 585 590

Leu Gln Val Lys Pro Gly Glu Ser Leu Glu Val Ile Gln Thr Thr Asp
595 600 605

Asp Thr Lys Val Leu Cys Arg Asn Glu Glu Gly Lys Tyr Gly Tyr Val
610 615 620

Leu Arg Ser Tyr Leu Ala Asp Asn Asp Gly Glu Ile Tyr Asp Asp Ile
625 630 635 640

Ala Asp Gly Cys Ile Tyr Asp Asn Asp
645

(2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

Met Ala Trp Pro Ser Arg Ser Lys Met Phe Thr Leu Leu Pro Val Leu
1 5 10 15

Cys Tyr Leu Trp Ser Leu Trp Leu Pro Gln Phe Ser Trp Ile Gln Glu
20 25 30

Leu Lys Ala Val Leu Arg Asp Asp Gly Leu Ile Ser Ala Val Ala Trp
35 40 45

Asn Ala Glu Phe Gln Thr Cys
50 55

(2) INFORMATION FOR SEQ ID NO: 203:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 267 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

Met Val Lys Val Thr Phe Asn Ser Ala Leu Ala Gln Lys Glu Ala Lys
1 5 10 15

Lys Asp Glu Pro Lys Ser Gly Glu Glu Ala Leu Ile Ile Pro Pro Asp
20 25 30

Ala Val Ala Val Asp Cys Lys Asp Pro Asp Asp Val Val Pro Val Gly
35 40 45

Gln Arg Arg Ala Trp Cys Trp Cys Met Cys Phe Gly Leu Ala Phe Met
50 55 60

Leu Ala Gly Val Ile Leu Gly Gly Ala Tyr Leu Tyr Lys Tyr Phe Ala
65 70 75 80

Leu Gln Pro Asp Asp Val Tyr Tyr Cys Gly Ile Lys Tyr Ile Lys Asp
85 90 95

Asp Val Ile Leu Asn Glu Pro Ser Ala Asp Ala Pro Ala Ala Leu Tyr
100 105 110

Gln Thr Ile Glu Glu Asn Ile Lys Ile Phe Glu Glu Glu Glu Val Glu
115 120 125

Phe Ile Ser Val Pro Val Pro Glu Phe Ala Asp Ser Asp Pro Ala Asn
130 135 140

Ile Val His Asp Phe Asn Lys Lys Leu Thr Ala Tyr Leu Asp Leu Asn
145 150 155 160

Leu Asp Lys Cys Tyr Val Ile Pro Leu Asn Thr Ser Ile Val Met Pro
165 170 175

Pro Arg Asn Leu Leu Glu Leu Leu Ile Asn Ile Lys Ala Gly Thr Tyr
180 185 190

Leu Pro Gln Ser Tyr Leu Ile His Glu His Met Val Ile Thr Asp Arg
195 200 205

Ile Glu Asn Ile Asp His Leu Gly Phe Phe Ile Tyr Arg Leu Cys His
210 215 220

Asp Lys Glu Thr Tyr Lys Leu Gln Arg Arg Glu Thr Ile Lys Gly Ile
225 230 235 240

Gln Lys Arg Glu Ala Ser Asn Cys Phe Ala Ile Arg His Phe Glu Asn

245 250 255

Lys Phe Ala Val Glu Thr Leu Ile Cys Ser Xaa
260 265

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 315 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

Met Asp Leu Arg Gln Phe Leu Met Cys Leu Ser Leu Cys Thr Ala Phe
1 5 10 15

Ala Leu Ser Lys Pro Thr Glu Lys Lys Asp Arg Val His His Glu Pro
20 25 30

Gln Leu Ser Asp Lys Val His Asn Asp Ala Gln Ser Phe Asp Tyr Asp
35 40 45

His Asp Ala Phe Leu Gly Ala Glu Glu Ala Lys Thr Phe Asp Gln Leu
50 55 60

Thr Pro Glu Glu Ser Lys Glu Arg Leu Gly Lys Ile Val Ser Lys Ile
65 70 75 80

Asp Gly Asp Lys Asp Gly Phe Val Thr Val Asp Glu Leu Lys Asp Trp
85 90 95

Ile Lys Phe Ala Gln Lys Arg Trp Ile Tyr Glu Asp Val Glu Arg Gln
100 105 110

Trp Lys Gly His Asp Leu Asn Glu Asp Gly Leu Val Ser Trp Glu Glu
115 120 125

Tyr Lys Asn Ala Thr Tyr Gly Tyr Val Leu Asp Asp Pro Asp Pro Asp
130 135 140

Asp Gly Phe Asn Tyr Lys Gln Met Met Val Arg Asp Glu Arg Arg Phe
145 150 155 160

Lys Met Ala Asp Lys Asp Gly Asp Leu Ile Ala Thr Lys Glu Glu Phe
165 170 175

Thr Ala Phe Leu His Pro Glu Glu Tyr Asp Tyr Met Lys Asp Ile Val
180 185 190

Val Gln Glu Thr Met Glu Asp Ile Asp Lys Asn Ala Asp Gly Phe Ile
195 200 205

Asp Leu Glu Glu Tyr Ile Gly Asp Met Tyr Ser His Asp Gly Asn Thr
210 215 220

Asp Glu Pro Glu Trp Val Lys Thr Glu Arg Glu Gln Phe Val Glu Phe
225 230 235 240

Arg Asp Lys Asn Arg Asp Gly Lys Met Asp Lys Glu Glu Thr Lys Asp
245 250 255

Trp Ile Leu Pro Ser Asp Tyr Asp His Ala Glu Ala Glu Ala Arg His
260 265 270

Leu Val Tyr Glu Ser Asp Gln Asn Lys Asp Gly Lys Leu Thr Lys Glu
275 280 285

Glu Ile Val Asp Lys Tyr Asp Leu Phe Val Gly Ser Gln Ala Thr Asp
290 295 300

Phe Gly Glu Ala Leu Val Arg His Asp Glu Phe
305 310 315

(2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

Met Phe Asp Ala Val Leu Ile Leu Leu Leu Ile Pro Leu Lys Asp Lys
1 5 10 15

Leu Val Asp Pro Ile Leu Arg Arg His Gly Leu Leu Pro Ser Ser Leu
20 25 30

Lys Arg Ile Ala Val Gly Met Phe Phe Val Met Cys Ser Ala Phe Ala
35 40 45

Ala Gly Ile Leu Glu Ser Lys Arg Leu Asn Leu Val Lys Glu Lys Thr
50 55 60

Ile Asn Gln Thr Ile Gly Asn Val Val Tyr His Ala Ala Asp Leu Ser
65 70 75 80

Leu Trp Trp Gln Val Pro Gln Tyr Leu Leu Ile Gly Ile Ser Glu Ile
85 90 95

Phe Ala Ser Ile Ala Gly Leu Glu Phe Ala Tyr Ser Ala Ala Pro Lys
100 105 110

Ser Met Gln Ser Ala Ile Met Gly Leu Phe Phe Phe Phe Ser Gly Val
115 120 125

Gly Ser Phe Val Gly Ser Gly Leu Leu Ala Leu Val Ser Ile Lys Ala
130 135 140

Ile Gly Trp Met Ser Ser His Thr Asp Phe Gly Asn Ile Asn Gly Cys
145 150 155 160

Tyr Leu Asn Tyr Tyr Phe Phe Leu Leu Ala Ala Ile Gln Gly Ala Thr
165 170 175

Leu Leu Leu Phe Leu Ile Ile Ser Val Lys Tyr Asp His His Arg Asp
180 185 190

His Gln Arg Ser Arg Ala Asn Gly Val Pro Thr Ser Arg Arg Ala
195 200 205

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 196 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

Met Arg Ser Arg Ile Arg Glu Phe Asp Ser Ser Thr Leu Asn Glu Ser
1 5 10 15

Val Arg Asn Thr Ile Met Arg Asp Leu Lys Ala Val Gly Lys Lys Phe
20 25 30

Met His Val Leu Tyr Pro Arg Lys Ser Asn Thr Leu Leu Arg Asp Trp
35 40 45

Asp Leu Trp Gly Pro Leu Ile Leu Cys Val Thr Leu Ala Leu Met Leu
50 55 60

Gln Arg Asp Ser Ala Asp Ser Glu Lys Asp Gly Gly Pro Gln Phe Ala
65 70 75 80

Glu Val Phe Val Ile Val Trp Phe Gly Ala Val Thr Ile Thr Leu Asn
85 90 95

Ser Lys Leu Leu Gly Gly Asn Ile Ser Phe Phe Gln Ser Leu Cys Val
100 105 110

Leu Gly Tyr Cys Ile Leu Pro Leu Thr Val Ala Met Leu Ile Cys Arg
115 120 125

Leu Val Leu Leu Ala Asp Pro Gly Pro Val Asn Phe Met Val Arg Leu
130 135 140

Phe Val Val Ile Val Met Phe Ala Trp Ser Ile Val Ala Ser Thr Ala
145 150 155 160

Phe Leu Ala Asp Ser Gln Pro Pro Asn Arg Arg Ala Leu Ala Val Tyr
165 170 175

Pro Val Phe Leu Phe Tyr Phe Val Ile Ser Trp Met Ile Leu Thr Phe
180 185 190

Thr Pro Gln Xaa
195

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 331 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

```
Met Ala Lys Asp Gln Ala Val Glu Asn Ile Leu Val Ser Pro Val Val
1               5                   10                  15

Val Ala Ser Ser Leu Gly Leu Val Ser Leu Gly Gly Lys Ala Thr Thr
            20                  25                  30

Ala Ser Gln Ala Lys Ala Val Leu Ser Ala Glu Gln Leu Arg Asp Glu
        35                  40                  45

Glu Val His Ala Gly Leu Gly Glu Leu Leu Arg Ser Leu Ser Asn Ser
    50                  55                  60

Thr Ala Arg Asn Val Thr Trp Lys Leu Gly Ser Arg Leu Tyr Gly Pro
65                  70                  75                  80

Ser Ser Val Ser Phe Ala Asp Asp Phe Val Arg Ser Ser Lys Gln His
                85                  90                  95

Tyr Asn Cys Glu His Ser Lys Ile Asn Phe Arg Asp Lys Arg Ser Ala
            100                 105                 110

Leu Gln Ser Ile Asn Glu Trp Ala Ala Gln Thr Thr Asp Gly Lys Leu
        115                 120                 125

Pro Glu Val Thr Lys Asp Val Glu Arg Thr Asp Gly Ala Leu Leu Val
    130                 135                 140

Asn Ala Met Phe Phe Lys Pro His Trp Asp Glu Lys Phe His His Lys
145                 150                 155                 160

Met Val Asp Asn Arg Gly Phe Met Val Thr Arg Ser Tyr Thr Val Gly
                165                 170                 175

Val Met Met Met His Arg Thr Gly Leu Tyr Asn Tyr Tyr Asp Asp Glu
            180                 185                 190

Lys Glu Lys Leu Gln Ile Val Glu Met Pro Leu Ala His Lys Leu Ser
        195                 200                 205

Ser Leu Ile Ile Leu Met Pro His His Val Glu Pro Leu Glu Arg Leu
    210                 215                 220

Glu Lys Leu Leu Thr Lys Glu Gln Leu Lys Ile Trp Met Gly Lys Met
225                 230                 235                 240

Gln Lys Lys Ala Val Ala Ile Ser Leu Pro Lys Gly Val Val Glu Val
                245                 250                 255

Thr His Asp Leu Gln Lys His Leu Ala Gly Leu Gly Leu Thr Glu Ala
            260                 265                 270

Ile Asp Lys Asn Lys Ala Asp Leu Ser Arg Met Ser Gly Lys Lys Asp
        275                 280                 285

Leu Tyr Leu Ala Ser Val Phe His Ala Thr Ala Phe Glu Leu Asp Thr
    290                 295                 300
```

Asp Gly Asn Pro Leu Thr Arg Ile Thr Gly Gly Gly Val Arg Thr Gln
305 310 315 320

Val Phe Tyr Ala Asp His Pro Phe Ile Ser Xaa
325 330

(2) INFORMATION FOR SEQ ID NO: 208:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

Met Cys Met Gln Leu Phe Gly Phe Leu Ala Phe Met Ile Phe Met Cys
1 5 10 15

Trp Val Gly Asp Val Tyr Pro Val Tyr Gln Pro Val Gly Pro Lys Gln
20 25 30

Tyr Pro Tyr Asn Asn Leu Tyr Leu Glu Arg Gly Gly Asp Pro Ser Lys
35 40 45

Glu Pro Glu Arg Val Val His Tyr Glu Ile
50 55

(2) INFORMATION FOR SEQ ID NO: 209:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

Met Asp Ala Leu Val Glu Asp Asp Ile Cys Ile Leu Asn His Glu Lys
1 5 10 15

Ala His Lys Arg Asp Thr Val Thr Pro Val Ser Ile Tyr Ser Gly Asp
20 25 30

Glu Ser Val Ala Ser His Phe Ala Leu Val Thr Ala Tyr Glu Asp Ile
35 40 45

Lys Lys Arg Leu Lys Asp Ser Glu Lys Glu Asn Ser Leu Leu Lys Lys
50 55 60

Arg Ile Arg Phe Leu Glu Glu Lys Leu Ile Ala Arg Phe Glu Glu Glu
65 70 75 80

Thr Ser Ser Val Gly Arg Glu Gln Val Asn Lys Ala Tyr His Ala Tyr
85 90 95

Arg Glu Val Cys Ile Asp Arg Asp Asn Leu Lys Ser Lys Leu Asp Lys
100 105 110

Met Asn Lys Asp Asn Ser Glu Ser Leu Lys Val Leu Asn Glu Gln Leu

115 120 125

Gln Ser Lys Glu Val Glu Leu Leu Gln Leu Arg Thr Glu Val Glu Thr
130 135 140

Gln Gln Val Met Arg Asn Leu Asn Pro Pro Ser Ser Asn Trp Glu Val
145 150 155 160

Glu Lys Leu Ser Cys Asp Leu Lys Ile His Gly Leu Glu Gln Glu Leu
165 170 175

Glu Leu Met Arg Lys Glu Cys Ser Asp Leu Lys Ile Glu Leu Gln Lys
180 185 190

Ala Lys Gln Thr Asp Pro Tyr Gln Glu Asp Asn Leu Lys Ser Arg Asp
195 200 205

Leu Gln Lys Leu Ser Ile Ser Ser Asp Asn Met Gln His Ala Tyr Trp
210 215 220

Glu Leu Lys Arg Glu Met Ser Asn Leu His Leu Val Thr Gln Val Gln
225 230 235 240

Ala Glu Leu Leu Arg Lys Leu Lys Thr Ser Thr Ala Ile Lys Lys Ala
245 250 255

Cys Ala Pro Val Gly Cys Ser Glu Asp Leu Gly Arg Asp Ser Thr Lys
260 265 270

Leu His Leu Met Asn Phe Thr Ala Thr Tyr Thr Arg His Pro Pro Leu
275 280 285

Leu Pro Asn Gly Lys Ala Leu Cys His Thr Thr Ser Ser Pro Leu Pro
290 295 300

Gly Asp Val Lys Val Leu Ser Glu Lys Ala Ile Leu Gln Ser Trp Thr
305 310 315 320

Asp Asn Glu Arg Ser Ile Pro Asn Asp Gly Thr Cys Phe Gln Glu His
325 330 335

Ser Ser Tyr Gly Arg Asn Ser Leu Glu Asp Asn Ser Trp Val Phe Pro
340 345 350

Ser Pro Pro Lys Ser Ser Glu Thr Ala Phe Gly Glu Thr Lys Thr Lys
355 360 365

Thr Leu Pro Leu Pro Asn Leu Pro Pro Leu His Tyr Leu Asp Gln His
370 375 380

Asn Gln Asn Cys Leu Tyr Lys Asn
385 390

(2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

```
Met His His His Thr Gln Leu Met Phe Ile Tyr Leu Phe Ile Tyr Leu
 1               5                  10                  15

Phe Ile Leu Gly Val Phe Phe Phe Phe Phe Xaa
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

```
Met Asn Cys Ile Leu Leu Leu Tyr Leu Leu Ile Pro Thr Ile Ser Ile
 1               5                  10                  15

Ser Val Val Pro Tyr Val Ala Leu Asn Ile Lys Tyr Ile Lys Glu Cys
            20                  25                  30

Thr Glu Asn Ser Phe Tyr Xaa
        35
```

(2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

```
Met Leu Leu His Leu Thr Ala Ala Phe Leu Gln Arg Ala Gln Phe Ser
 1               5                  10                  15

Thr Tyr Phe Pro Gly Tyr Phe Asp Gly Gln Tyr Trp Leu Trp Trp Val
            20                  25                  30

Phe Leu Val Leu Gly Phe Leu Leu Phe Leu Arg Gly Phe Ile Asn Tyr
        35                  40                  45

Ala Lys Val Arg Lys Met Pro Glu Thr Phe Ser Asn Leu Pro Arg Thr
    50                  55                  60

Arg Val Leu Phe Ile Tyr Xaa
65                  70
```

(2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

```
Met Leu Thr Phe Phe Met Ala Phe Leu Phe Asn Trp Ile Gly Phe Phe
 1               5                  10                  15

Leu Ser Phe Cys Leu Thr Thr Ser Ala Ala Gly Arg Tyr Gly Ala Ile
            20                  25                  30

Ser Gly Phe Gly Leu Ser Leu Ile Lys Trp Ile Leu Ile Val Arg Phe
        35                  40                  45

Ser Thr Tyr Phe Pro Ala Phe Met Asn Ser Leu Ser Arg Ser Lys Arg
    50                  55                  60

Thr Pro Ala Gly Ser Glu Ser Arg Cys Arg Thr Gln Arg Asn Asn His
65                  70                  75                  80

Leu Leu Xaa
```

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

```
Met Ser Lys Arg Ser Ala Ser Phe Ile Leu Leu Pro Leu Leu Phe Leu
 1               5                  10                  15

Lys Gly Ser Phe Ala Lys Leu Asn Ala Arg Ile Ser Asp Cys Leu Glu
            20                  25                  30

Glu Arg Tyr Cys His Asn Leu Trp Met Val Phe Gln Gly Cys Val Ile
        35                  40                  45

Thr Glu Leu His Leu Ser Arg Met Ser Lys Thr Leu Ser Ser Leu Cys
    50                  55                  60

Tyr Asp Phe Val Ile Asn Val Tyr Ile Phe Phe Lys Phe Leu Asp Ile
65                  70                  75                  80

Thr
```

(2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

```
Met Cys Ser Leu Phe Glu Ser Arg Phe Phe Cys Phe Val Leu Phe Ser
```

1 5 10 15

Glu Lys Ile Ile Gln Leu Cys Ala Ser Ile Ala Phe Leu Cys Phe Val
20 25 30

Lys His Val Pro Trp Pro Lys Trp Lys Arg Lys Cys Leu Ile Asn Ala
35 40 45

Phe

(2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 203 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

Met Thr Leu Arg Pro Ser Leu Leu Pro Leu His Leu Leu Leu Leu Leu
1 5 10 15

Leu Leu Ser Ala Ala Val Cys Arg Ala Glu Ala Gly Leu Glu Thr Glu
20 25 30

Ser Pro Val Arg Thr Leu Gln Val Glu Thr Leu Val Glu Pro Pro Glu
35 40 45

Pro Cys Ala Glu Pro Ala Ala Phe Gly Asp Thr Leu His Ile His Tyr
50 55 60

Thr Gly Ser Leu Val Asp Gly Arg Ile Ile Asp Thr Ser Leu Thr Arg
65 70 75 80

Asp Pro Leu Val Ile Glu Leu Gly Gln Lys Gln Val Ile Pro Gly Leu
85 90 95

Glu Gln Ser Leu Leu Asp Met Cys Val Gly Glu Lys Arg Arg Ala Ile
100 105 110

Ile Pro Ser His Leu Ala Tyr Gly Lys Arg Gly Phe Pro Pro Ser Val
115 120 125

Pro Ala Asp Ala Val Val Gln Tyr Asp Val Glu Leu Ile Ala Leu Ile
130 135 140

Arg Ala Asn Tyr Trp Leu Lys Leu Val Lys Gly Ile Leu Pro Leu Val
145 150 155 160

Gly Met Ala Met Val Pro Pro Ser Trp Ala Ser Leu Gly Ile Thr Tyr
165 170 175

Thr Glu Arg Pro Ile Asp Pro Lys Ser Pro Lys Arg Ser Ser Arg Lys
180 185 190

Arg Asn Glu Thr Arg Ala Lys Arg Asn Asn Lys
195 200

(2) INFORMATION FOR SEQ ID NO: 217:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

Met Lys Thr Leu Met Thr Ile Cys Pro Gly Thr Val Leu Leu Val Phe
1   5   10   15

Ser Ile Ser Leu Trp Ile Ile Ala Ala Trp Thr Val Arg Val Cys Glu
            20   25   30

Ser Pro Glu Ser Pro Ala Gln Pro Ser Gly Ser Ser Leu Pro Ala Trp
        35   40   45

Tyr His Asp Gln Gln Asp Val Thr Ser Asn Phe Leu Gly Ala Met Trp
    50   55   60

Leu Ile Ser Ile Thr Phe Leu Ser Ile Gly Tyr Gly Asp Met Val Pro
65   70   75   80

His Thr Tyr Cys Gly Lys Gly Val Cys Leu Leu Thr Gly Ile Met Gly
                85   90   95

Ala Gly Cys Thr Ala Leu Val Val Ala Val Val Ala Arg Lys Leu Glu
            100   105   110

Leu Thr Lys Ala Glu Lys His Val His Xaa Phe Met Met Asp Thr Gln
        115   120   125

Leu Thr Lys Arg Ile Lys Asn Xaa Ala Ala Asn Val Leu Xaa Glu Thr
    130   135   140

Trp Leu Ile Tyr Lys His Thr Lys Leu Leu Lys Lys Ile Asp His Ala
145   150   155   160

Lys Val Arg Asn Thr Arg Gly Ser Ser Ser Lys Tyr Pro Pro Val Glu
                165   170   175

Glu Arg Gln Asp Gly Thr Glu Glu Ala Glu
            180   185

(2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

Met Lys Phe Leu Ala Val Leu Val Leu Leu Gly Val Ser Ile Phe Leu
1   5   10   15

Val Ser Ala Gln Asn Pro Thr Thr Ala Ala Pro Ala Asp Thr Tyr Pro

20 25 30

Ala Thr Gly Pro Ala Asp Asp Glu Ala Pro Asp Ala Glu Thr Thr Ala
35 40 45

Ala Ala Thr Thr Ala Thr Thr Ala Ala Pro Thr Thr Ala Thr Thr Ala
50 55 60

Ala Ser Thr Thr Ala Arg Lys Asp Ile Pro Val Leu Pro Lys Trp Val
65 70 75 80

Gly Asp Leu Pro Asn Gly Arg Val Cys Pro
85 90

(2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 139 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

Met Ser Ser Ala Ala Ala Asp His Trp Ala Trp Leu Leu Val Leu Ser
1 5 10 15

Phe Val Phe Gly Cys Asn Val Leu Arg Ile Leu Leu Pro Ser Phe Ser
20 25 30

Ser Phe Met Ser Arg Val Leu Gln Lys Asp Ala Glu Gln Glu Ser Gln
35 40 45

Met Arg Ala Glu Ile Gln Asp Met Lys Gln Glu Leu Ser Thr Val Asn
50 55 60

Met Met Asp Glu Phe Ala Arg Tyr Ala Arg Leu Glu Arg Lys Ile Asn
65 70 75 80

Lys Met Thr Asp Lys Leu Lys Thr His Val Lys Ala Arg Thr Ala Gln
85 90 95

Leu Ala Lys Ile Lys Trp Val Ile Ser Val Ala Phe Tyr Val Leu Gln
100 105 110

Ala Ala Leu Met Ile Ser Leu Ile Trp Lys Tyr Tyr Ser Val Pro Val
115 120 125

Ala Val Val Pro Ser Lys Trp Ile Thr Leu Xaa
130 135

(2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

Met Ser Ser Ala Ala Ala Asp His Trp Ala Trp Leu Leu Val Leu Ser
1 5 10 15

Phe Val Phe Gly Cys Asn Val Leu Arg Ile Leu Leu Pro Ser Phe Ser
20 25 30

Ser Phe Met Ser Arg Val Leu Gln Lys Asp Ala Asp Arg Ser His Arg
35 40 45

(2) INFORMATION FOR SEQ ID NO: 221:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

Met Thr Ala Pro Leu Pro Pro Leu Ser Gly Leu Ala Leu Phe Leu Ile
1 5 10 15

Val Phe Phe Ser Leu Gly Val Phe Cys Ile Cys His Ser His Trp Tyr
20 25 30

His Thr Leu Gln Gln Met Ala Gly Thr Glu Pro Lys Ala Leu Leu Leu
35 40 45

Ser Pro Pro Ala Ala Thr Thr Phe Val Thr Val Thr His Glu Val Trp
50 55 60

Lys Glu Gln Ala Leu Ala
65 70

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

Met Thr Cys Ser Val Ala Leu Leu Leu Ile Leu Gly Leu Arg Cys Ser
1 5 10 15

Gly Val Arg Pro Gly Leu Val Gly Glu Gly His Asn Pro Ser Leu Leu
20 25 30

Val Cys Leu Leu Leu Lys Asp Ser Arg Thr Asn Gln Gly Ser Cys Pro
35 40 45

Gly Gly Pro Trp Ser Glu Arg Asp Ile Glu Ser Val Thr Ser Asp Asn
50 55 60

Cys Glu Ala Thr Leu Gly Tyr Arg Asn His Ser Leu Pro Ser Asn Tyr
65 70 75 80

Tyr Asn Ser

(2) INFORMATION FOR SEQ ID NO: 223:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

Met Leu Thr Arg Ser Leu Lys Thr Leu Pro Ser Ala Cys Thr Ala Phe
1 5 10 15

Leu Leu Leu Phe Phe Leu Phe Ser Ser Gly Asp Pro Glu Leu Ser Cys
20 25 30

Ser Cys Thr Leu Arg Thr Gln Ser Ser Trp Ser
35 40

(2) INFORMATION FOR SEQ ID NO: 224:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 184 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

Met Trp Arg Pro Ser Val Leu Leu Leu Leu Leu Leu Leu Arg His Gly
1 5 10 15

Ala Gln Gly Lys Pro Ser Pro Asp Ala Gly Pro His Gly Gln Gly Arg
20 25 30

Val His Gln Ala Ala Pro Leu Ser Asp Ala Pro His Asp Asp Ala His
35 40 45

Gly Asn Phe Gln Tyr Asp His Glu Ala Phe Leu Gly Arg Glu Val Ala
50 55 60

Lys Glu Phe Asp Gln Leu Thr Pro Glu Glu Ser Gln Ala Arg Leu Gly
65 70 75 80

Arg Ile Val Asp Arg Met Asp Arg Ala Gly Asp Gly Asp Gly Trp Val
85 90 95

Ser Leu Ala Glu Leu Arg Ala Trp Ile Ala His Thr Gln Gln Arg His
100 105 110

Ile Arg Asp Ser Val Ser Ala Ala Trp Asp Thr Tyr Asp Thr Asp Arg
115 120 125

Asp Gly Arg Val Gly Trp Glu Glu Leu Arg Asn Xaa Thr Tyr Gly His

130 135 140

Xaa Xaa Pro Xaa Glu Glu Phe His Asp Val Glu Asp Ala Glu Thr Tyr
145 150 155 160

Lys Lys Met Leu Xaa Arg Asp Glu Arg Arg Phe Arg Val Ala Asp Gln
165 170 175

Asp Gly Asp Ser Met Ala Thr Arg
180

(2) INFORMATION FOR SEQ ID NO: 225:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

Met Trp Leu Phe Ile Leu Leu Ser Leu Ala Leu Ile Ser Asp Ala Met
1 5 10 15

Val Met Asp Glu Lys Val Lys Arg Ser Leu Cys Trp Thr Arg Leu Leu
20 25 30

Pro Ser Ala Thr Thr Met Pro Xaa Thr Arg Ile Thr Pro Asn Thr Gly
35 40 45

Ala Glu Xaa Ile Ser Val Xaa Thr Ala Thr Ser Ser Pro Ser Pro Leu
50 55 60

Thr Ala Pro Ile Met Trp Pro
65 70

(2) INFORMATION FOR SEQ ID NO: 226:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

Met His Val Phe Val Leu Glu Ile Phe Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

Met Ala Val Ala Thr Leu Ala Ser Glu Thr Leu Pro Leu Leu Ala Leu

```
 1               5                   10                  15
Thr Phe Ile Thr Asp Asn Ser Leu Val Ala Ala Gly His Asp Cys Phe
            20                  25                  30
Pro Val Leu Phe Thr Tyr Asp Ala Ala Ala Gly Met Leu Ser Phe Gly
        35                  40                  45
Gly Arg Leu Asp Val Pro Lys Gln Ser Ser Gln Arg Gly Leu Thr Ala
    50                  55                  60
Arg Glu Arg Phe Gln Asn Leu Asp Lys Lys Ala Ser Ser Glu Gly Gly
65                  70                  75                  80
Thr Ala Ala Gly Ala Gly Leu Asp Ser Leu His Lys Asn Ser Val Ser
                85                  90                  95
Gln Ile Ser Val Leu Ser Gly Gly Lys Ala Lys Cys Ser Gln Phe Cys
            100                 105                 110
Thr Thr Gly Met Asp Gly Gly Met Ser Ile Trp Asp Val Lys Ser Leu
        115                 120                 125
Glu Ser Ala Leu Lys Asp Leu Lys Ile Lys
    130                 135
```

(2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

```
Leu Gly Ser Leu Ser Thr Ala Pro Ser Ser Ala Leu Pro Thr Leu Gly
 1               5                   10                  15
Ala Arg Arg Thr Arg Ser Lys
            20
```

(2) INFORMATION FOR SEQ ID NO: 229:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

```
Met Thr Tyr Phe Ser Gly Leu Leu Val Ile Leu Ala Phe Ala Ala Trp
 1               5                   10                  15
Val Ala Leu Ala Glu Gly Leu Gly Val Ala Val Tyr Ala Ala Ala Val
            20                  25                  30
Leu Leu Gly Ala Gly Cys Ala Thr Ile Leu Val Thr Ser Leu Ala Met
        35                  40                  45
```

Thr Ala Asp Leu Ile Gly Pro His Thr Asn Ser Gly Ala Phe Val Tyr
50 55 60

Gly Ser Met Ser Phe Leu Asp Lys Val Ala Asn Gly Leu Ala Val Met
65 70 75 80

Ala Ile Gln Ser Leu His Pro Cys Pro Ser Glu Leu Cys Cys Arg Ala
85 90 95

Cys Val Ser Phe Tyr His Trp Ala Met Val Ala Val Thr Gly Gly Val
100 105 110

Gly Val Ala Ala Ala Leu Cys Leu Cys Ser Leu Leu Leu Trp Pro Thr
115 120 125

Arg Leu Arg Arg Xaa
130

(2) INFORMATION FOR SEQ ID NO: 230:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

Gly Lys Pro Thr Gly Lys Ser Leu Pro Leu Met Trp Met Ile Leu Met
1 5 10 15

Gln Pro Ile Ile Met Ile Ser Met Met Ser Asn Gly
20 25

(2) INFORMATION FOR SEQ ID NO: 231:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

Met Gln Gly Lys Phe Met Lys Val Gln Val Tyr Arg Phe Leu Lys Tyr
1 5 10 15

Leu Leu Met Leu Leu Cys Met Phe Val Asn Arg Gly Met Ser Lys Asp
20 25 30

Ser Thr Lys Lys Pro Gly Gln Glu Lys Leu Lys Val Ser Leu Gly Ser
35 40 45

Ile Leu Asn Met Lys Ser Gln Arg Pro Leu Ser Trp Cys
50 55 60

(2) INFORMATION FOR SEQ ID NO: 232:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

```
Met Met Glu Arg Ser Met Met Ile Leu Leu Met Ala Ala Ser Met Thr
  1               5                  10                  15

Met Thr Ser Thr Gln Leu Trp Ser Phe Cys Cys Val His
             20                  25
```

(2) INFORMATION FOR SEQ ID NO: 233:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

```
Met Trp Tyr Gln Leu Ala Lys Glu Glu Pro Gly Val Gly Ala Cys Ala
  1               5                  10                  15

Leu Asp
```

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

```
Leu Xaa
  1
```

(2) INFORMATION FOR SEQ ID NO: 235:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

```
Met Leu Ile Cys Arg Leu Val Leu Leu Ala Asp Pro Gly Pro Val Asn
  1               5                  10                  15

Phe Met Val Arg Leu Phe Val Val Ile Val Met Phe Ala Trp Ser Ile
             20                  25                  30

Val Ala Ser Thr Ala Phe Leu Ala Asp Ser Gln Pro Pro Asn Arg Arg
         35                  40                  45
```

Ala Leu Ala Val Tyr Pro Val Phe Leu Phe Tyr Phe Val Ile Ser Trp
50 55 60

Met Ile Leu Thr Phe Thr Pro Gln
65 70

(2) INFORMATION FOR SEQ ID NO: 236:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

Met Arg Ser Leu Leu Leu Leu Ser Ala Phe Cys Leu Leu Glu Ala Ala
1 5 10 15

Leu Ala Ala Glu Val Lys Lys Pro Ala Ala Ala Ala Ala Pro Gly Thr
20 25 30

Ala Glu Lys Leu Ser Pro Lys Ala Ala Thr Leu Ala Glu Arg Xaa Pro
35 40 45

Ala Trp Pro Ser Ala Cys Thr Arg Pro Trp Pro Arg Thr Arg Gln Trp
50 55 60

Arg Thr Ser Trp Cys His Pro Trp Trp Trp Pro Arg Arg Trp Gly Ser
65 70 75 80

Cys Arg Trp Ala Ala Arg Arg Pro Arg Arg Arg Arg Pro Arg Gln Cys
85 90 95

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

Met Arg Ser Leu Leu Leu Leu Ser Ala Phe Cys Leu Leu Glu Ala Ala
1 5 10 15

Leu Ala Ala Glu Val Lys Lys Pro Ala Ala Ala Ala Ala Pro Gly Thr
20 25 30

Ala Glu Lys Leu Ser Pro Lys Ala Ala Thr Leu Ala Glu Arg Lys Arg
35 40 45

Pro Gly Leu Gln Leu Val Pro Gly His Gly Gln Gly Pro Gly Ser Gly
50 55 60

Glu His Pro Gly Val Thr Arg Gly Gly Gly Leu Val Ala Gly Ala Arg
65 70 75 80

Val Ala Gly Arg Gln Gly Asp His Gly Val Ala Gly Gln Gly Ser Ala
85 90 95

Glu Arg Arg Ala Ala Ala Arg Arg Gly Gly Ala Arg Arg Pro Gly Arg
100 105 110

Ala Ala Ala Leu Thr Gln Gln Leu His Gly Ala Gln Arg Asp Leu Glu
115 120 125

Ala Gly Gln Pro Thr Val Arg Thr Gln Leu Ser Glu Leu Arg Xaa
130 135 140

(2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 142 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

Met Arg Ser Leu Leu Leu Leu Ser Ala Phe Cys Leu Leu Glu Ala Ala
1 5 10 15

Leu Ala Ala Glu Val Lys Lys Pro Ala Ala Ala Ala Ala Pro Gly Thr
20 25 30

Ala Glu Lys Leu Ser Pro Lys Ala Ala Thr Leu Ala Glu Arg Xaa Arg
35 40 45

Pro Gly Leu Gln Leu Val Pro Gly His Gly Gln Gly Pro Gly Ser Gly
50 55 60

Glu His Pro Gly Val Thr Arg Gly Gly Gly Leu Val Ala Gly Ala Arg
65 70 75 80

Val Ala Gly Arg Gln Gly Asp His Gly Val Ala Gly Gln Gly Ser Ala
85 90 95

Glu Arg Arg Ala Ala Ala Arg Arg Gly Gly Ala Arg Arg Pro Gly Arg
100 105 110

Ala Ala Ala Leu Thr Gln Gln Leu Xaa Gly Ala Gln Arg Asp Leu Glu
115 120 125

Ala Gly Gln Pro Thr Val Arg Thr Gln Leu Ser Glu Leu Arg
130 135 140

(2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

Asp Pro Glu Ala Ala Asp Ser Gly Glu Pro Gln Asn Lys Arg Thr Pro
1   5   10   15

Asp Leu Pro Glu Glu Glu Tyr Val Lys Glu Glu Ile Gln Glu Asn Glu
            20   25   30

Glu Ala Val Lys Lys Met Leu Val Glu Ala Thr Arg Glu Phe Glu Glu
        35   40   45

Val Val Val Asp Glu Ser
    50

(2) INFORMATION FOR SEQ ID NO: 240:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

Gln Lys Leu Lys Arg Lys Ala Glu Glu Asp Pro Glu Ala Ala Asp Ser
1   5   10   15

Gly Glu Pro Gln Asn Lys Arg Thr Pro Asp Leu Pro Glu Glu Glu Tyr
            20   25   30

Val Lys Glu Glu Ile Gln Glu Asn Glu Glu Ala Val Lys Lys Met Leu
        35   40   45

Val Glu Ala Thr Arg Glu Phe Glu Glu Val Val Val Asp Glu Ser
    50   55   60

(2) INFORMATION FOR SEQ ID NO: 241:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 113 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

Lys Ala Met Glu Lys Ser Ser Leu Thr Gln His Ser Trp Gln Ser Leu
1   5   10   15

Lys Asp Arg Tyr Leu Lys His Leu Arg Gly Gln Glu His Lys Tyr Leu
            20   25   30

Leu Gly Asp Ala Pro Val Ser Pro Ser Ser Gln Lys Leu Lys Arg Lys
        35   40   45

Ala Glu Glu Asp Pro Glu Ala Ala Asp Ser Gly Glu Pro Gln Asn Lys
    50   55   60

Arg Thr Pro Asp Leu Pro Glu Glu Glu Tyr Val Lys Glu Glu Ile Gln
65   70   75   80

Glu Asn Glu Glu Ala Val Lys Lys Met Leu Val Glu Ala Thr Arg Glu
85 90 95

Phe Glu Glu Val Val Val Asp Glu Ser Pro Pro Asp Phe Glu Ile His
100 105 110

Ile

(2) INFORMATION FOR SEQ ID NO: 242:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 148 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

Leu Pro Ser Tyr Asp Glu Ala Glu Arg Thr Lys Ala Glu Ala Thr Ile
1 5 10 15

Pro Leu Val Pro Gly Arg Asp Glu Asp Phe Val Gly Arg Asp Asp Phe
20 25 30

Asp Asp Ala Asp Gln Leu Arg Ile Gly Asn Asp Gly Ile Phe Met Leu
35 40 45

Thr Phe Phe Met Ala Phe Leu Phe Asn Trp Ile Gly Phe Phe Leu Ser
50 55 60

Phe Cys Leu Thr Thr Ser Ala Ala Gly Arg Tyr Gly Ala Ile Ser Gly
65 70 75 80

Phe Gly Leu Ser Leu Ile Lys Trp Ile Leu Ile Val Arg Phe Ser Thr
85 90 95

Tyr Phe Pro Gly Tyr Phe Asp Gly Gln Tyr Trp Leu Trp Trp Val Phe
100 105 110

Leu Val Leu Gly Phe Leu Leu Phe Leu Arg Gly Phe Ile Asn Tyr Ala
115 120 125

Lys Val Arg Lys Met Pro Glu Thr Phe Ser Asn Leu Pro Arg Thr Arg
130 135 140

Val Leu Phe Ile
145

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

Ala Gly Arg Tyr Gly Ala Ile Ser Gly Phe Gly Leu Ser Leu Ile Lys
1 5 10 15

Trp Ile Leu Ile Val Arg Phe Ser
20

(2) INFORMATION FOR SEQ ID NO: 244:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

Met Lys His Leu Ser Ala Trp Asn Phe Thr Lys Leu Thr Phe Leu Gln
1 5 10 15

Leu Trp Glu Ile Phe Glu Gly Ser Val Glu Asn Cys Gln Thr Leu Thr
20 25 30

Ser Tyr Ser Lys Leu Gln Ile Lys Tyr Thr Phe Ser Arg Gly Ser Thr
35 40 45

Phe Tyr Ile
50

(2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 213 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

Phe Ser Ser Asp Phe Arg Thr Ser Pro Trp Glu Ser Arg Arg Val Glu
1 5 10 15

Ser Lys Ala Thr Ser Ala Arg Cys Gly Leu Trp Gly Ser Gly Pro Arg
20 25 30

Arg Arg Pro Ala Ser Gly Met Phe Arg Gly Leu Ser Ser Trp Leu Gly
35 40 45

Leu Gln Gln Pro Val Ala Gly Gly Gly Gln Pro Asn Gly Asp Ala Pro
50 55 60

Pro Glu Gln Pro Ser Glu Thr Val Ala Glu Ser Ala Glu Glu Glu Leu
65 70 75 80

Gln Gln Ala Gly Asp Gln Glu Leu Leu His Gln Ala Lys Asp Phe Gly
85 90 95

Asn Tyr Leu Phe Asn Phe Ala Ser Ala Ala Thr Lys Lys Ile Thr Glu
100 105 110

Ser Val Ala Glu Thr Ala Gln Thr Ile Lys Lys Ser Val Glu Glu Gly

115 120 125

Lys Ile Asp Gly Ile Ile Asp Lys Thr Ile Ile Gly Asp Phe Gln Lys
130 135 140

Glu Gln Lys Lys Phe Val Glu Glu Gln His Thr Lys Lys Ser Glu Ala
145 150 155 160

Ala Val Pro Pro Trp Val Asp Thr Asn Asp Glu Glu Thr Ile Gln Gln
165 170 175

Gln Ile Leu Ala Leu Ser Ala Asp Lys Arg Asn Phe Leu Arg Asp Pro
180 185 190

Pro Ala Gly Val Gln Phe Asn Phe Asp Phe Asp Gln Met Tyr Pro Val
195 200 205

Ala Leu Val Met Leu
210

(2) INFORMATION FOR SEQ ID NO: 246:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

Met Arg Phe Ala Leu Val Pro Lys Leu Val Lys Glu Glu Val Phe Trp
1 5 10 15

Arg Asn Tyr Phe Tyr Arg Val Ser Leu Ile Lys Gln Ser Ala Gln Leu
20 25 30

Thr Ala Leu Ala Ala Gln Gln Gln Ala Ala Gly Lys Gly Gly Glu Glu
35 40 45

Gln

(2) INFORMATION FOR SEQ ID NO: 247:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 76 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

Ser Thr Ser Pro Gly Val Ser Glu Phe Val Ser Asp Ala Phe Asp Ala
1 5 10 15

Cys Asn Leu Asn Gln Glu Asp Leu Arg Lys Glu Met Glu Gln Leu Val
20 25 30

Leu Asp Lys Lys Gln Glu Glu Thr Ala Val Leu Glu Glu Asp Ser Ala
35 40 45

Asp Trp Glu Lys Glu Leu Gln Gln Glu Leu Gln Glu Tyr Glu Val Val
50 55 60

Thr Glu Ser Glu Lys Arg Asp Glu Asn Trp Asp Lys
65 70 75

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 62 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

Ser Pro Trp Glu Ser Arg Arg Val Glu Ser Lys Ala Thr Ser Ala Arg
1 5 10 15

Cys Gly Leu Trp Gly Ser Gly Pro Arg Arg Arg Pro Ala Ser Gly Met
20 25 30

Phe Arg Gly Leu Ser Ser Trp Leu Gly Leu Gln Gln Pro Val Ala Gly
35 40 45

Gly Gly Gln Pro Asn Gly Asp Ala Pro Pro Glu Gln Pro Ser
50 55 60

(2) INFORMATION FOR SEQ ID NO: 249:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

Pro Val Ala Gly Gly Gly Gln Pro Asn Gly Asp Ala Pro Pro Glu Gln
1 5 10 15

Pro Ser Glu Thr Val Ala Glu Ser Ala Glu Glu Glu Leu Gln Gln Ala
20 25 30

Gly Asp Gln Glu Leu Leu His Gln Ala Lys Asp Phe Gly Asn Tyr Leu
35 40 45

Phe Asn Phe Ala Ser Ala Ala Thr Lys Lys Ile Thr Glu Ser Val Ala
50 55 60

Glu
65

(2) INFORMATION FOR SEQ ID NO: 250:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

```
Phe Gln Lys Glu Gln Lys Lys Phe Val Glu Glu Gln His Thr Lys Lys
 1               5                  10                  15

Ser Glu Ala Ala Val Pro Pro Trp Val Asp Thr Asn Asp Glu Glu Thr
            20                  25                  30

Ile Gln Gln Gln Ile Leu Ala Leu Ser Ala Asp Lys Arg Asn Phe Leu
        35                  40                  45

Arg Asp Pro Pro Ala Gly Val Gln Phe Asn Phe Asp Phe Asp Gln Met
    50                  55                  60

Tyr Pro Val Ala Leu Val Met Leu
65                  70
```

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

```
Pro Phe Ile Cys Val Ala Arg Asn Pro Val Ser Arg Asn Phe Ser Ser
 1               5                  10                  15

Pro Ile Leu Ala Arg Lys Leu Cys Glu Gly Ala Ala
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

```
Lys Glu Asp Pro Ala Asn Thr Val Tyr Ser Thr Val Glu Ile Pro Lys
 1               5                  10                  15

Lys Met Glu Asn Pro His Ser Leu Leu Thr Met Pro Asp Thr Pro Arg
            20                  25                  30

Leu
```

(2) INFORMATION FOR SEQ ID NO: 253:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 227 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

Ala Ser Ala Val Leu Leu Asp Leu Pro Asn Ser Gly Gly Glu Ala Gln
1 5 10 15

Ala Lys Lys Leu Gly Asn Asn Cys Val Phe Ala Pro Ala Asp Val Thr
20 25 30

Ser Glu Lys Asp Val Gln Thr Ala Leu Ala Leu Ala Lys Gly Lys Phe
35 40 45

Gly Arg Val Asp Val Ala Val Asn Cys Ala Gly Ile Ala Val Ala Ser
50 55 60

Lys Thr Tyr Asn Leu Lys Lys Gly Gln Thr His Thr Leu Glu Asp Phe
65 70 75 80

Gln Arg Val Leu Asp Val Asn Leu Met Gly Thr Phe Asn Val Ile Arg
85 90 95

Leu Val Ala Gly Glu Met Gly Gln Asn Glu Pro Asp Gln Gly Gly Gln
100 105 110

Arg Gly Val Ile Ile Asn Thr Ala Ser Val Ala Ala Phe Glu Gly Gln
115 120 125

Val Gly Gln Ala Ala Tyr Ser Ala Ser Lys Gly Gly Ile Val Gly Met
130 135 140

Thr Leu Pro Ile Ala Arg Asp Leu Ala Pro Ile Gly Ile Arg Val Met
145 150 155 160

Thr Ile Ala Pro Gly Leu Phe Gly Thr Pro Leu Leu Thr Ser Leu Pro
165 170 175

Glu Lys Val Cys Asn Phe Leu Ala Ser Gln Val Pro Phe Pro Ser Arg
180 185 190

Leu Gly Asp Pro Ala Glu Tyr Ala His Leu Val Gln Ala Ile Ile Glu
195 200 205

Asn Pro Phe Leu Asn Gly Glu Val Ile Arg Leu Asp Gly Ala Ile Arg
210 215 220

Met Gln Pro
225

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

Ser Val Ala Ala Phe Glu Gly Gln Val Gly Gln Ala Ala Tyr Ser Ala

```
1               5                   10                  15

Ser Lys Gly Gly Ile Val Gly Met Thr Leu Pro Ile Ala
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 255:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

```
Ala Arg Arg Ser Gly Ala Glu Leu Ala Trp Asp Tyr Leu Cys Arg Trp
1               5                   10                  15

Ala Gln Lys His Lys Asn Trp Arg Phe Gln Lys Thr Arg Gln Thr Trp
            20                  25                  30

Leu Leu Leu His Met Tyr Asp Ser Asp Lys Val Pro Asp Glu His Phe
        35                  40                  45

Ser Thr Leu Leu Ala Tyr Leu Glu Gly Leu Gln Gly Arg
    50                  55                  60
```

(2) INFORMATION FOR SEQ ID NO: 256:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

```
His Pro Ile Glu Trp Ala Ile Asn Ala Ala Thr Leu Ser Gln Phe Tyr
1               5                   10                  15

Ile Asn Lys Leu Cys Phe
            20
```

(2) INFORMATION FOR SEQ ID NO: 257:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

```
Cys Trp Ile Lys Tyr Cys Leu Thr Leu Met Gln Asn Ala Gln Leu Ser
1               5                   10                  15

Met Gln Asp Asn Ile Gly
            20
```

(2) INFORMATION FOR SEQ ID NO: 258:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

Lys Val Ser Tyr Leu Arg Pro Leu Asp Phe Glu Glu Ala Arg Glu Leu
1 5 10 15

Phe Leu Leu Gly Gln His Tyr Val Phe
20 25

(2) INFORMATION FOR SEQ ID NO: 259:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

Met Glu Arg Arg Cys Lys Met His Lys Arg Xaa Ile Ala Met Leu Glu
1 5 10 15

Pro Leu Thr Val Asp Leu Asn Pro Gln
20 25

(2) INFORMATION FOR SEQ ID NO: 260:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

Ser His Ile Val Lys Lys Ile Asn Asn Leu Asn Lys Ser Ala Leu Lys
1 5 10 15

Tyr Tyr Gln Leu Phe Leu Asp
20

(2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

Phe Thr His Leu Ser Thr Cys Leu Leu Ser Leu Leu Leu Val Arg Met
1 5 10 15

Ser Gly Phe Leu Leu Leu Ala Arg Ala Ser Pro Ser Ile Cys Ala Leu

20 25 30

Asp Ser Ser Cys Phe Val Gln Glu Tyr Cys Ser Ser Tyr Ser Ser Ser
35 40 45

Cys Phe Leu His Gln His Phe Pro Ser Leu Leu Asp His Leu Cys Gln
50 55 60

(2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

Phe Leu Leu Leu Ala Arg Ala Ser Pro Ser Ile Cys Ala Leu Asp Ser
1 5 10 15

Ser Cys Phe Val Gln Glu Tyr
20

(2) INFORMATION FOR SEQ ID NO: 263:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

Pro Asp Gly Arg Val Thr Asn Ile Pro Gln Gly Met Val Thr Asp Gln
1 5 10 15

Phe Gly Met Ile Gly Leu Leu Thr Phe Ile Arg Ala Ala Glu Thr Asp
20 25 30

Pro Gly Met Val His Leu Ala Leu Gly Ser Asp Leu Thr Thr Leu Gly
35 40 45

Leu Asn Leu Asn Ser
50

(2) INFORMATION FOR SEQ ID NO: 264:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

Glu Asp Leu Leu Phe Tyr Leu Tyr Tyr Met Asn Gly Gly Asp Val Leu

1 5 10 15

Gln Leu Leu Ala Ala Val Glu Leu Phe Asn Arg Asp Trp Arg Tyr His
20 25 30

Lys Glu Glu Arg Val Trp Ile Thr Arg
35 40

(2) INFORMATION FOR SEQ ID NO: 265:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

Val His Leu Ala Leu Gly Ser Asp Leu Thr Thr Leu Gly Leu Asn Leu
1 5 10 15

Asn Ser Pro Glu Asn Leu Tyr Pro
20

(2) INFORMATION FOR SEQ ID NO: 266:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

His Asn Glu Asp Phe Pro Ala Leu Pro Gly Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 267:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

Gly Arg Ile Ile Asp Thr Ser Leu Thr Arg Asp Pro Leu Val Ile Glu
1 5 10 15

Leu Gly Gln Lys Gln Val Ile Pro Gly Leu Glu Gln Ser Leu Leu Asp
20 25 30

Met Cys Val Gly Glu Lys Arg Arg Ala Ile Ile Pro Ser His Leu Ala
35 40 45

Tyr Gly Lys Arg Gly Phe Pro Pro Ser Val Pro Ala Asp Ala Val Val
50 55 60

Gln Tyr Asp Val Glu Leu Ile Ala Leu Ile Arg

65 70 75

(2) INFORMATION FOR SEQ ID NO: 268:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

Ile His Tyr Thr Gly Ser Leu Val Asp Gly Arg Ile Ile Asp Thr Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 269:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

Cys Glu Ser Pro Glu Ser Pro Ala Gln Pro Ser Gly Ser Ser Leu Pro
1 5 10 15

Ala Trp Tyr His
20

(2) INFORMATION FOR SEQ ID NO: 270:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 95 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

Glu Glu Ala Gly Ala Gly Arg Arg Cys Ser His Gly Gly Ala Arg Pro
1 5 10 15

Ala Gly Leu Gly Asn Glu Gly Leu Gly Leu Gly Gly Asp Pro Asp His
20 25 30

Thr Asp Thr Gly Ser Arg Ser Lys Gln Arg Ile Asn Asn Trp Lys Glu
35 40 45

Ser Lys His Lys Val Ile Met Ala Ser Ala Ser Ala Arg Gly Asn Gln
50 55 60

Asp Lys Asp Ala His Phe Pro Pro Pro Ser Lys Gln Ser Leu Leu Phe
65 70 75 80

Cys Pro Lys Ser Lys Leu His Ile His Arg Ala Glu Ile Ser Lys

85 90 95

(2) INFORMATION FOR SEQ ID NO: 271:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

Ser Lys Gln Arg Ile Asn Asn Trp Lys Glu Ser Lys His Lys Val Ile
1 5 10 15

Met Ala Ser Ala Ser Ala Arg
20

(2) INFORMATION FOR SEQ ID NO: 272:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

Leu Phe His Trp Ala Cys Leu Asn Glu Arg Ala Ala Gln Leu Pro Arg
1 5 10 15

Asn Thr Ala Xaa Ala Gly Tyr Gln Cys Pro Ser Cys Asn Gly Pro Ser
20 25 30

(2) INFORMATION FOR SEQ ID NO: 273:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 185 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

Phe Tyr Ile Tyr Tyr Arg Pro Thr Asp Ser Asp Asn Asp Ser Asp Tyr
1 5 10 15

Lys Lys Asp Met Val Glu Gly Asp Lys Tyr Trp His Ser Ile Ser His
20 25 30

Leu Gln Pro Glu Thr Ser Tyr Asp Ile Lys Met Gln Cys Phe Asn Glu
35 40 45

Gly Gly Glu Ser Glu Phe Ser Asn Val Met Ile Cys Glu Thr Lys Ala
50 55 60

Arg Lys Ser Ser Gly Gln Pro Gly Arg Leu Pro Pro Pro Thr Leu Ala

65 70 75 80

Pro Pro Gln Pro Pro Leu Pro Glu Thr Ile Glu Arg Pro Val Gly Thr
85 90 95

Gly Ala Met Val Ala Arg Ser Ser Asp Leu Pro Tyr Leu Ile Val Gly
100 105 110

Val Val Leu Gly Ser Ile Val Leu Ile Ile Val Thr Phe Ile Pro Phe
115 120 125

Cys Leu Trp Arg Ala Trp Ser Lys Gln Lys His Thr Thr Asp Leu Gly
130 135 140

Phe Pro Arg Ser Ala Leu Pro Pro Ser Cys Pro Tyr Thr Met Val Pro
145 150 155 160

Leu Gly Gly Leu Pro Gly His Gln Ala Val Asp Ser Pro Thr Ser Val
165 170 175

Ala Ser Val Asp Gly Pro Val Leu Met
180 185

(2) INFORMATION FOR SEQ ID NO: 274:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

Tyr Ile Tyr Tyr Arg Pro Thr Asp Ser Asp Asn Asp Ser Asp Tyr Lys
1 5 10 15

Lys Asp Met Val Glu Gly Asp Lys Tyr Trp His Ser Ile Ser His Leu
20 25 30

Gln Pro Glu Thr Ser Tyr Asp Ile Lys Met Gln Cys Phe Asn Glu Gly
35 40 45

Gly Glu Ser Glu Phe Ser Asn Val Met Ile Cys Glu Thr Lys Ala Arg
50 55 60

Lys Ser
65

(2) INFORMATION FOR SEQ ID NO: 275:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

Asn Val Arg Ala Leu Leu His Arg Met Pro Glu Pro Pro Lys Ile Asn
1 5 10 15

Thr Ala Lys Phe Asn Asn Asn Lys Arg Lys Asn Leu Ser Leu
20 25 30

(2) INFORMATION FOR SEQ ID NO: 276:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 185 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

Asn Thr Asn Gln Arg Glu Ala Leu Gln Tyr Ala Lys Asn Phe Gln Pro
1 5 10 15

Phe Ala Leu Asn His Gln Lys Asp Ile Gln Val Leu Met Gly Ser Leu
20 25 30

Val Tyr Leu Arg Gln Gly Ile Glu Asn Ser Pro Tyr Val His Leu Leu
35 40 45

Asp Ala Asn Gln Trp Ala Asp Ile Cys Asp Ile Phe Thr Arg Asp Ala
50 55 60

Cys Ala Leu Leu Gly Leu Ser Val Glu Ser Pro Leu Ser Val Ser Phe
65 70 75 80

Ser Ala Gly Cys Val Ala Leu Pro Ala Leu Ile Asn Ile Lys Ala Val
85 90 95

Ile Glu Gln Arg Gln Cys Thr Gly Val Trp Asn Gln Lys Asp Glu Leu
100 105 110

Pro Ile Glu Val Asp Leu Gly Lys Lys Cys Trp Tyr His Ser Ile Phe
115 120 125

Ala Cys Pro Ile Leu Arg Gln Gln Thr Thr Asp Asn Asn Pro Pro Met
130 135 140

Lys Leu Val Cys Gly His Ile Ile Ser Arg Asp Ala Leu Asn Lys Met
145 150 155 160

Phe Asn Gly Ser Lys Leu Lys Cys Pro Tyr Cys Pro Met Glu Gln Ser
165 170 175

Pro Gly Asp Ala Lys Gln Ile Phe Phe
180 185

(2) INFORMATION FOR SEQ ID NO: 277:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

Ser Tyr Leu Ser Ala Cys Phe Ala Gly Cys Asn Ser Thr Asn Leu Thr
1 5 10 15

Gly Cys Ala Cys Leu Thr Thr Val Pro Ala Glu Asn Ala Thr Val Val
20 25 30

Pro Gly Lys Cys Pro Ser Pro Gly Cys Gln Glu Ala Phe Leu Thr Phe
35 40 45

Leu Cys Val Met Cys Ile Cys Ser Leu Ile Gly Ala Met Ala Arg His
50 55 60

Pro
65

(2) INFORMATION FOR SEQ ID NO: 278:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 84 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

Pro Ser Val Ile Ile Leu Ile Arg Thr Val Ser Pro Glu Leu Lys Ser
1 5 10 15

Tyr Ala Leu Gly Val Leu Phe Leu Leu Leu Arg Leu Leu Gly Phe Ile
20 25 30

Pro Pro Pro Leu Ile Phe Gly Ala Gly Ile Asp Ser Thr Cys Leu Phe
35 40 45

Trp Ser Thr Phe Cys Gly Glu Gln Gly Ala Cys Val Leu Tyr Asp Asn
50 55 60

Val Val Tyr Arg Tyr Leu Tyr Val Ser Ile Ala Ile Ala Leu Lys Ser
65 70 75 80

Phe Ala Phe Ile

(2) INFORMATION FOR SEQ ID NO: 279:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 182 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

Gln Ser Leu Phe Thr Arg Phe Val Arg Val Gly Val Pro Thr Val Asp
1 5 10 15

Leu Asp Ala Gln Gly Arg Ala Arg Ala Ser Leu Cys Xaa Xaa Tyr Asn
20 25 30

Trp Arg Tyr Lys Asn Leu Gly Asn Leu Pro His Val Gln Leu Leu Pro

35 40 45

Glu Phe Ser Thr Ala Asn Ala Gly Leu Leu Tyr Asp Phe Gln Leu Ile
50 55 60

Asn Val Glu Asp Phe Gln Gly Val Gly Glu Ser Glu Pro Asn Pro Tyr
65 70 75 80

Phe Tyr Gln Asn Leu Gly Glu Ala Glu Tyr Val Val Ala Leu Phe Met
85 90 95

Tyr Met Cys Leu Leu Gly Tyr Pro Ala Asp Lys Ile Ser Ile Leu Thr
100 105 110

Thr Tyr Asn Gly Gln Lys His Leu Ile Arg Asp Ile Ile Asn Arg Arg
115 120 125

Cys Gly Asn Asn Pro Leu Ile Gly Arg Pro Asn Lys Val Thr Thr Val
130 135 140

Asp Arg Phe Gln Gly Gln Gln Asn Asp Tyr Ile Leu Leu Ser Leu Val
145 150 155 160

Arg Thr Arg Ala Val Gly His Leu Arg Asp Val Arg Arg Leu Val Val
165 170 175

Ala Met Ser Arg Ala Arg
180

(2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 77 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

Leu Val Lys Glu Ala Lys Ile Ile Ala Met Thr Cys Thr His Ala Ala
1 5 10 15

Leu Lys Arg His Asp Leu Val Lys Leu Gly Phe Lys Tyr Asp Asn Ile
20 25 30

Leu Met Glu Glu Ala Ala Gln Ile Leu Glu Ile Glu Thr Phe Ile Pro
35 40 45

Leu Leu Leu Gln Asn Pro Gln Asp Gly Phe Ser Arg Leu Lys Arg Trp
50 55 60

Ile Met Ile Gly Asp His His Gln Leu Pro Pro Val Ile
65 70 75

(2) INFORMATION FOR SEQ ID NO: 281:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

Asp Thr Tyr Pro Asn Glu Glu Lys Gln Gln Glu Arg Val Phe Pro Xaa
1 5 10 15

Xaa Ser Ala Met Val Asn Asn Gly Ser Leu Ser Tyr Asp His Glu Arg
20 25 30

Asp Gly Arg Pro Thr Glu Leu Gly Gly Cys Xaa Ala Ile Val Arg Asn
35 40 45

Leu His Tyr Asp Thr Phe Leu Val Ile Arg Tyr Val Lys Arg His Leu
50 55 60

Thr Ile Met Met Asp Ile Asp Gly Lys His Glu Trp Arg Asp Cys Ile
65 70 75 80

Glu Val Pro Gly Val Arg Leu Pro Arg Gly Tyr Tyr Phe Gly Thr Ser
85 90 95

Ser Ile Thr Gly Asp Leu Ser Asp Asn His Asp Val Ile Ser Leu Lys
100 105 110

Leu Phe Glu Leu Thr Val Glu Arg Thr Pro Glu Glu Glu
115 120 125

(2) INFORMATION FOR SEQ ID NO: 282:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 85 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

Leu Lys Arg Glu His Ser Leu Ser Lys Pro Tyr Gln Gly Val Gly Thr
1 5 10 15

Gly Ser Ser Ser Leu Trp Asn Leu Met Gly Asn Ala Met Val Met Thr
20 25 30

Gln Tyr Ile Arg Leu Thr Pro Asp Met Gln Ser Lys Gln Gly Ala Leu
35 40 45

Trp Asn Arg Val Pro Cys Phe Leu Arg Asp Trp Glu Leu Gln Val His
50 55 60

Phe Lys Ile His Gly Gln Gly Lys Lys Asn Leu His Gly Asp Gly Leu
65 70 75 80

Ala Ile Trp Tyr Thr
85

(2) INFORMATION FOR SEQ ID NO: 283:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

Pro Gly Thr Leu Gln Cys Ser Ala Leu His His Asp Pro Gly Cys Ala
1 5 10 15

Asn Cys Ser Arg Phe Cys Arg Asp Cys Ser Pro Pro Ala Cys Gln Cys
20 25 30

(2) INFORMATION FOR SEQ ID NO: 284:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

Phe Leu Tyr Asp Val Leu Met Xaa His Glu Ala Val Met Arg Thr His
1 5 10 15

Gln Ile Gln Leu Pro Asp Pro Glu Phe Pro Ser
20 25

(2) INFORMATION FOR SEQ ID NO: 285:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

Gly Trp Tyr Trp Cys Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 286:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

Met Lys Val Gly Ala Arg Ile Arg Val Lys Met Ser Val Asn Lys Ala
1 5 10 15

His Pro Val Val Ser Thr His Trp Arg Trp Pro Ala Glu Trp Pro Gln
20 25 30

Met Phe Leu His Leu Ala Gln Glu Pro Arg Thr Glu Val Lys Ser Arg
35 40 45

Pro Leu Gly Leu Ala Gly Phe Ile Arg Gln Asp Ser Lys Thr Arg Lys
50 55 60

Pro Leu Glu Gln Glu Thr Ile Met Ser Ala Ala Asp Thr Ala Leu Trp
65 70 75 80

Pro Tyr Gly His Gly Asn Arg Glu His Gln Glu Asn Glu Leu Gln Lys
85 90 95

Tyr Leu Gln Tyr Lys Asp Met His Leu Leu Asp Ser Gly Gln Ser Leu
100 105 110

Gly His Thr His Thr Leu Gln Gly Ser His Asn Leu Thr Ala Leu Asn
115 120 125

Ile

(2) INFORMATION FOR SEQ ID NO: 287:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

Ser Leu His Lys Asn Ser Val Ser Gln Ile Ser Val Leu Ser Gly Gly
1 5 10 15

Lys Ala Lys Cys Ser Gln Phe Cys Thr Thr Gly Met Asp Gly Gly Met
20 25 30

Ser Ile Trp Asp Val Lys Ser Leu Glu Ser Ala Leu Lys Asp Leu Lys
35 40 45

Ile

(2) INFORMATION FOR SEQ ID NO: 288:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

Glu Ala Ser Lys Ser Ser His Ala Gly Leu Asp Leu Phe Ser Val Ala
1 5 10 15

Ala Cys His Arg Phe
20

(2) INFORMATION FOR SEQ ID NO: 289:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

Tyr Met Gly Lys Gly Ser Met Thr Gly Leu Ala Leu Lys His Met Phe
1 5 10 15

Glu Arg Ser Phe Thr
20

(2) INFORMATION FOR SEQ ID NO: 290:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

Val Thr Gly Ile Ile Asp Ser Leu Thr Ile Ser Pro Lys Ala Ala Arg
1 5 10 15

Val Gly Leu Leu Gln Tyr Ser Thr Gln Val His
20 25

(2) INFORMATION FOR SEQ ID NO: 291:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

Thr Glu Phe Thr Leu Arg Asn Phe Asn Ser Ala Lys Asp Met Lys Lys
1 5 10 15

Ala Val Ala His Met Lys Tyr Met
20

(2) INFORMATION FOR SEQ ID NO: 292:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

Gly Lys Gly Ser Met Thr Gly Leu Ala Leu Lys His Met Phe Glu Arg
1 5 10 15

```
Ser Phe Thr Gln Gly Glu Gly Ala Arg Pro Phe
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 293:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

```
Ser Thr Arg Val Pro Arg Ala Ala Ile Val Phe Thr Asp Gly Arg Ala
  1               5                  10                  15

Gln Asp Asp Val Ser Glu Trp Ala Ser Lys Ala Lys Ala Asn Gly Ile
            20                  25                  30

Thr Met Tyr Ala Val Gly Val Gly Lys Ala Ile Glu
        35                  40
```

(2) INFORMATION FOR SEQ ID NO: 294:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

```
Glu Glu Leu Gln Glu Ile Ala Ser Glu Pro Thr Asn Lys His Leu Phe
  1               5                  10                  15

Tyr Ala Glu Asp Phe Ser Thr Met Asp Glu Ile Ser Glu Lys Leu Lys
            20                  25                  30

Lys Gly Ile Cys Glu Ala Leu Glu Asp Ser
        35                  40
```

(2) INFORMATION FOR SEQ ID NO: 295:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

```
Thr Gln Arg Leu Glu Glu Met Thr Gln Arg Met
  1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 296:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

Pro Gln Gly Cys Pro Glu Gln Pro Leu His
1 5 10

(2) INFORMATION FOR SEQ ID NO: 297:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

Arg Cys Lys Lys Cys Thr Glu Gly Pro Ile Asp Leu Val Phe Val Ile
1 5 10 15

Asp Gly Ser Lys Ser Leu Gly Glu Glu Asn Phe Glu Val Val Lys Gln
20 25 30

Phe

(2) INFORMATION FOR SEQ ID NO: 298:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

Met Ala Ala Leu Leu Leu Arg His Val Gly Arg His Cys Leu Arg Ala
1 5 10 15

His Phe Ser Pro Gln Leu Cys Ile Arg Asn Ala Val Pro Leu Gly Thr
20 25 30

Thr Ala Lys Glu Glu Met Glu Arg Phe Trp Asn Lys Asn Ile Gly Ser
35 40 45

Asn Arg Pro Leu Ser Pro His Ile Thr Ile Tyr Ser
50 55 60

(2) INFORMATION FOR SEQ ID NO: 299:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

Val Phe Pro Leu Met Tyr His Thr Trp Asn Gly Ile Arg His Leu Met
1 5 10 15

Trp Asp Leu Gly Lys Gly Leu Lys Ile Pro Gln Leu Tyr Gln Ser Gly
20 25 30

(2) INFORMATION FOR SEQ ID NO: 300:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

Met Ala Ala Leu Leu Leu Arg His Val Gly Arg His Cys Leu Arg Ala
1 5 10 15

His

(2) INFORMATION FOR SEQ ID NO: 301:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

Val Lys Ser Leu Cys Leu Gly Pro Ala Leu Ile His Thr Ala Lys Phe
1 5 10 15

Ala Leu

(2) INFORMATION FOR SEQ ID NO: 302:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

Val Phe Pro Leu Met Tyr His Thr Trp Asn Gly Ile Arg His Leu Met
1 5 10 15

Trp Asp Leu Gly Lys Gly Leu
20

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

```
Arg Val Trp Asp Val Arg Pro Phe Ala Pro Lys Glu Arg Cys Val Lys
  1               5                  10                  15

Ile Phe Gln Gly Asn Val
             20
```

(2) INFORMATION FOR SEQ ID NO: 304:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

```
His Asn Phe Glu Lys Asn Leu Leu Arg Cys Ser Trp Ser Pro Asp Gly
  1               5                  10                  15

Ser Lys Ile Ala Ala Gly Ser Ala Asp Arg Phe Val Tyr Val
             20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

```
Trp Asp Thr Thr Ser Arg Arg Ile Leu Tyr Lys Leu Pro Gly His Ala
  1               5                  10                  15

Gly Ser Ile Asn Glu Val Ala Phe His Pro Asp Glu Pro Ile
             20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

```
Val Arg Gly Arg Thr Val Leu Arg Pro Gly Leu Asp Ala Glu Pro Glu
  1               5                  10                  15

Leu Ser Pro Glu
             20
```

(2) INFORMATION FOR SEQ ID NO: 307:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

Glu Gln Arg Val Leu Glu Arg Lys Leu Lys Lys Glu Arg Lys Lys Glu
1 5 10 15

Glu Arg Gln

(2) INFORMATION FOR SEQ ID NO: 308:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

Arg Leu Arg Glu Ala Gly Leu Val Ala Gln His Pro Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO: 309:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

Gly Arg Ile Pro Ala Pro Ala Pro Ser Val Pro Ala Gly Pro Asp Ser
1 5 10 15

Arg

(2) INFORMATION FOR SEQ ID NO: 310:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

Thr Gly Cys Val Leu Val Leu Ser Arg Asn Phe Val Gln Tyr Ala Cys
1 5 10 15

Phe Gly Leu Phe Gly Ile Ile Ala Leu Gln Thr Ile Ala Tyr Ser Ile
20 25 30

Leu Trp Asp Leu Lys Phe Leu Met Arg Asn
35 40

(2) INFORMATION FOR SEQ ID NO: 311:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

Ser Arg Ser Glu Gly Lys Ser Met Phe Ala Gly Val Pro Thr Met Arg
1 5 10 15

Glu Ser Ser Pro Lys Gln Tyr Met Gln Leu Gly Gly Arg Val Leu Leu
20 25 30

Val Leu Met Phe Met Thr Leu Leu His Phe Asp Ala Ser Phe Phe Ser
35 40 45

Ile Val Gln Asn Ile Val Gly
50 55

(2) INFORMATION FOR SEQ ID NO: 312:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

Gly Thr Ala Glu Asp Phe Ala Asp Gln Phe Leu Arg Val Thr Lys Gln
1 5 10 15

Tyr Leu Pro His Val Ala Arg Leu Cys Leu Ile Ser Thr Phe Leu Glu
20 25 30

Asp Gly Ile Arg Met Trp Phe Gln Trp Ser Glu Gln Arg Asp Tyr Ile
35 40 45

Asp Thr Thr Trp Asn Cys Gly Tyr Leu Leu Ala Ser
50 55 60

(2) INFORMATION FOR SEQ ID NO: 313:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

Ala Ser Phe Leu Leu Ser Arg Thr Ser Trp Gly Thr Ala Leu Met Ile
1 5 10 15

Leu

(2) INFORMATION FOR SEQ ID NO: 314:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Leu Met Arg Asn Glu Ser Arg Ser
1 5

(2) INFORMATION FOR SEQ ID NO: 315:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

Ala Ser Phe Leu Leu Ser Arg Thr Ser Trp Gly Thr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO: 316:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

Phe Ile Ser Phe Ala Asn Ser Arg Ser Ser Glu Asp Thr Lys Gln Met
1 5 10 15

Met Ser Ser Phe
20

(2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

Asp Pro Arg Arg Pro Asn Lys Val Leu Arg Tyr Lys Pro Pro Pro Ser
1 5 10 15

Glu Cys Asn Pro Ala Leu Asp Asp Pro Thr Pro
20 25

(2) INFORMATION FOR SEQ ID NO: 318:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

```
Asp Tyr Met Asn Leu Leu Gly Met Ile Phe Ser Met Cys Gly Leu Met
  1               5                  10                  15

Leu Lys Leu Lys Trp Cys Ala Trp Val Ala Val Tyr Cys Ser
             20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 319:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

```
Met Leu Ser Ile Ser Ala Val Val Met Ser Tyr Leu Gln Asn Pro Gln
  1               5                  10                  15

Pro Met Thr Pro Pro Trp
             20
```

(2) INFORMATION FOR SEQ ID NO: 320:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

```
Ala Ala Gly Asp Gly Asp Val Lys Leu Gly Thr Leu Gly Ser Gly Ser
  1               5                  10                  15

Glu Ser Ser Asn Asp Gly Gly Ser Glu Ser Pro Gly Asp Ala Gly Ala
             20                  25                  30

Ala Ala Xaa Gly Gly Gly Trp Ala Ala Ala Ala Leu Ala Leu Leu Thr
         35                  40                  45

Gly Gly Gly Glu
     50
```

(2) INFORMATION FOR SEQ ID NO: 321:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 177 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

Ala Ala Asp Asn Tyr Gly Ile Pro Arg Ala Cys Arg Asn Ser Ala Arg
1 5 10 15

Ser Tyr Gly Ala Ala Trp Leu Leu Leu Xaa Pro Ala Gly Ser Ser Arg
20 25 30

Val Glu Pro Thr Gln Asp Ile Ser Ile Ser Asp Gln Leu Gly Gly Gln
35 40 45

Asp Val Pro Val Phe Arg Asn Leu Ser Leu Leu Val Val Gly Val Gly
50 55 60

Ala Val Phe Ser Leu Leu Phe His Leu Gly Thr Arg Glu Arg Arg Arg
65 70 75 80

Pro His Ala Xaa Glu Pro Gly Glu His Thr Pro Leu Leu Ala Pro Ala
85 90 95

Thr Ala Gln Pro Leu Leu Leu Trp Lys His Trp Leu Arg Glu Xaa Ala
100 105 110

Phe Tyr Gln Val Gly Ile Leu Tyr Met Thr Thr Arg Leu Ile Val Asn
115 120 125

Leu Ser Gln Thr Tyr Met Ala Met Tyr Leu Thr Tyr Ser Leu His Leu
130 135 140

Pro Lys Lys Phe Ile Ala Thr Ile Pro Leu Val Met Tyr Leu Ser Gly
145 150 155 160

Phe Leu Ser Ser Phe Leu Met Lys Pro Ile Asn Lys Cys Ile Gly Arg
165 170 175

Asn

(2) INFORMATION FOR SEQ ID NO: 322:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

Arg Ile Thr Asp Asn Pro Glu Gly Lys Trp Leu Gly Arg Thr Ala Arg
1 5 10 15

Gly Ser Tyr Gly Tyr Ile Lys Thr Thr Ala Val Glu Ile Xaa Tyr Asp
20 25 30

Ser Leu Lys Leu Lys Lys Asp Ser Leu Gly Ala Pro Ser Arg Pro Ile
35 40 45

Glu Asp Asp Gln Glu Val Tyr Asp Asp Val Ala Glu Gln Asp Asp Ile

50 55 60

Ser Ser His Ser Gln Ser Gly Ser Gly Gly Ile Phe Pro Pro Pro Pro
65 70 75 80

Asp Asp Asp Ile Tyr Asp Gly Ile Glu Glu Glu Asp Ala Asp Asp Gly
85 90 95

Phe Pro Ala Pro Pro Lys Gln Leu Asp Met Gly Asp Glu Val Tyr Asp
100 105 110

Asp Val Asp Thr Ser Asp Phe Pro Val Ser Ser Ala Glu Met Ser Gln
115 120 125

Gly Thr Asn Val Gly Lys Ala Lys Thr Glu Glu Lys Asp Leu Lys Lys
130 135 140

Leu Lys Lys Gln Xaa Lys Glu Xaa Lys Asp Phe Arg Lys Lys Phe Lys
145 150 155 160

Tyr Asp Gly Glu Ile Arg Val Leu Tyr Ser Thr Lys Val Thr Thr Ser
165 170 175

Ile Thr Ser Lys Lys Trp Gly Thr Arg Asp Leu Gln Val Lys Pro Gly
180 185 190

Glu Ser Leu Glu Val Ile Gln Thr Thr Asp Asp Thr Lys Val Leu Cys
195 200 205

Arg Asn Glu Glu Gly Lys Tyr Gly Tyr Val Leu Arg Ser Tyr Leu Ala
210 215 220

Asp Asn Asp Gly Glu Ile Tyr Asp Asp Ile Ala Asp Gly Cys Ile Tyr
225 230 235 240

Asp Asn Asp

(2) INFORMATION FOR SEQ ID NO: 323:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 106 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

Ser Met Ser Ala Leu Thr Arg Leu Ala Ser Phe Ala Arg Val Gly Gly
1 5 10 15

Arg Leu Phe Arg Ser Gly Cys Ala Arg Thr Ala Gly Asp Gly Gly Val
20 25 30

Arg His Ala Gly Gly Gly Val His Ile Glu Pro Arg Tyr Arg Gln Phe
35 40 45

Pro Gln Leu Thr Arg Ser Gln Val Phe Gln Ser Glu Phe Phe Ser Gly
50 55 60

Leu Met Trp Phe Trp Ile Leu Trp Arg Phe Trp His Asp Ser Glu Gly

65 70 75 80

Val Leu Gly His Phe Pro Tyr Pro Asp Pro Ser Gln Trp Thr Asp Glu
85 90 95

Glu Leu Gly Ile Pro Pro Asp Asp Glu Asp
100 105

Applicant's or agent's file reference number	2004PCT	International application	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 73, line N/A

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit March 7, 1997	Accession Number 97923

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer Virginia L. Riley	Authorized officer

Applicant's or agent's file reference number	Z004PCT	International application	Unassigned PCT/US 98

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 73 , line N/A .

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	May 22, 1997	Accession Number	209071

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications. e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer Virginia L ...	Authorized officer

Applicant's or agent's file reference number	Z004PCT	International application	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis)*

A. The indications made below relate to the microorganism referred to in the description on page 73 , line N/A

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution *(including postal code and country)*

10801 University Boulevard
Manassas. Virginia 20110-2209
United States of America

Date of deposit February 25. 1998	Accession Number 209641

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications. e.g.. "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L [signature]	Authorized officer:

Applicant's or agent's file reference number	Z004PCT	International application No. Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis)*

A. The indications made below relate to the microorganism referred to in the description on page 75 , line N/A

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	July 24, 1997	Accession Number	209179

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Applicant's or agent's file reference number	Z004PCT	International application	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 77, line N/A.

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	March 7, 1997	Accession Number	97924

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) — This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L [signature]	Authorized officer

Applicant's or agent's file reference number	Z004PCT	International application No.	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 80 , line N/A .

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	March 13, 1997	Accession Number	97958

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer Virginia L. [signature]	Authorized officer

Applicant's or agent's file reference number	Z004PCT	International application	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 80, line N/A.

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	Accession Number
May 22, 1997	209072

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) — This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L Lilly	Authorized officer

Applicant's or agent's file reference number:	Z004PCT	International application No.	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis)*

A. The indications made below relate to the microorganism referred to in the description on page 80, line N/A

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution *(including postal code and country)*

10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit September 4, 1997	Accession Number 209235

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications, e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on
Authorized officer: Virginia L Miley	Authorized officer

Applicant's or agent's file reference number	Z004PC1	International application	Unassigned

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 84 , line N/A

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution: American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	August 28, 1997	Accession Number	209226

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L Lily	Authorized officer:

Applicant's or agent's file reference number Z004PCT	International application No. Unassigned PCT/US 98/

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 84, line N/A

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit March 13, 1997	Accession Number 97957

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L. [signature]	Authorized officer

Applicant's or agent's file reference number	Z004PCT	International application	Unassigned PCT/US 98

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis)*

A. The indications made below relate to the microorganism referred to in the description on page 84, line N/A.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit May 22, 1997	Accession Number 209073

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer Virginia L Nelson	Authorized officer

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity:

(f) a polynucleotide which is a variant of SEQ ID NO:X;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:X;

(h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y:

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(g) a variant of SEQ ID NO:Y;

(h) an allelic variant of SEQ ID NO:Y; or

(i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 22.

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C12N5/10 C12N1/21 C07K14/47 C07K16/18
C12Q1/68 G01N33/50 G01N33/53 G01N33/68 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	L. HILLIER ET AL.: "The WashU-Merck EST project" EMBL SEQUENCE DATABASE, 2 July 1995, HEIDELBERG, FRG, XP002068365 yl87a06.r1 Homo sapiens cDNA clone 44938 5'; Accession no. H08241; ---	1-3, 7-11,21
X	L. HILLIER ET AL.: "The "WashU-Merck EST project" EMBL SEQUENCE DATABASE, 26 August 1995, HEIDELBERG, FRG, XP002068366 ym94e01.r1 Homo sapiens cDNA clone 166584 5', Accession no. R88485; ---	1-3, 7-11,21
A	WO 97 07198 A (GENETICS INSTITUT) 27 February 1997 see the whole document ---	1-23

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 June 1998	16. 09. 1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	HORNIG H.

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 04097 A (GENETICS INST) 6 February 1997 see the whole document ---	1-23
A	US 5 536 637 A (JACOBS KENNETH) 16 July 1996 see the whole document ---	1-23
A	JACOBS K ET AL: "A novel method for isolating eukaryotic cDNA clones encoding secreted proteins." KEYSTONE SYMPOSIUM ON DENDRITIC CELLS: ANTIGEN PRESENTING CELLS OF T AND B LYMPHOCYTES, TAOS, NEW MEXICO, USA, MARCH 10-16, 1995. JOURNAL OF CELLULAR BIOCHEMISTRY SUPPLEMENT 0 (21A). 1995. 19. ISSN: 0733-1959, XP002027246 abstract no. C1-207 see abstract ---	1-23
A	WO 90 14432 A (GENETICS INST) 29 November 1990 see the whole document ---	1-23
A	WO 96 17925 A (IMMUNEX CORP) 13 June 1996 see the whole document -----	1-23

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claim 17 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see further information sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos..

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see further information sheet

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1. Claims: (1-23) partially

-An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from SEQ ID no. 11; wherein said polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein encoding the sequence SEQ ID no.125 or the polypeptide encoded by the cDNA sequence included in ATCC Deposit nos: 97923/209071 , which is hybridizable to SEQ ID no.11; a recombinant vector comprising said isolated nucleic acid molecule; a method of making a recombinant host cell comprising said isolated nucleic acid molecule; a recombinant host cell comprising said vector; an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence SEQ ID no. 125; an isolated antibody that binds specifically to said isolated polyeptitde; a recombinant host cell that expresses said isolated polypeptide; a method of making said polypeptide; a method for preventing, treating or ameliorating a medical condition which comprises administering to a mammalian subject a therapeutically effective amount of said polypeptide; a method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject using said polynucleotide and/or polypeptide sequence; a method for identifying a binding partner to said polypeptide; a gene corresponding to the cDNA sequence of SEQ ID no.11; a method for identifying an activity in a biological assay, by using the expression of SEQ ID no. 125;

Inventions 2 to 87. Claims: (12-23) partially

-Idem as subject 1 but limited to gene nos. 2 to 87 respectively cDNA clone sequences HAGFY16/HBMCF37/HFLQB16 to HCEDO21. (Invention 2 is limited to SEQ ID nos. 12,98,99,126,212 and 213; Invention 3 is limited to SEQ ID nos.13 and 127; ....; Invention 87 is limited to SEQ ID nos.97 and 211;)

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9707198 A	27-02-97	US 5707829 A	13-01-98
		AU 6712396 A	18-02-97
		AU 6768596 A	12-03-97
		EP 0839196 A	06-05-98
		EP 0851875 A	08-07-98
		WO 9704097 A	06-02-97
WO 9704097 A	06-02-97	US 5707829 A	13-01-98
		AU 6712396 A	18-02-97
		EP 0839196 A	06-05-98
		AU 6768596 A	12-03-97
		EP 0851875 A	08-07-98
		WO 9707198 A	27-02-97
US 5536637 A	16-07-96	US 5712116 A	27-01-98
WO 9014432 A	29-11-90	US 5580753 A	03-12-96
		AT 147436 T	15-01-97
		AU 637620 B	03-06-93
		AU 5928990 A	18-12-90
		CA 2056997 A	24-11-90
		DE 69029657 D	20-02-97
		DK 473724 T	14-04-97
		EP 0473724 A	11-03-92
		ES 2099096 T	16-05-97
		JP 4506006 T	22-10-92
		US 5734037 A	31-03-98
		US 5414071 A	09-05-95
WO 9617925 A	13-06-96	AU 4639396 A	26-06-96
		CA 2206488 A	13-06-96
		FI 972390 A	05-06-97
		NO 972455 A	06-08-97